NATIONAL ASSOCIATION OF BROADCASTERS

engineering handbook

9TH EDITION

NATIONAL
ASSOCIATION OF
BROADCASTERS

engineering handbook

9TH EDITION

Kenneth A. Sell
KZØX

NATIONAL ASSOCIATION OF BROADCASTERS engineering handbook

9TH EDITION

JERRY WHITAKER
Editor-in-Chief

JANET H. ELLIOTT
Managing Editor

COURTENAY SMITH BROWN
Technical Editor

NAB™
BROADCASTERS

This publication is not intended to be a substitute for the professional advice of an engineer or attorney. While the Editors have endeavored to include information current at the time of publication, statutes, rules, regulations and engineering knowledge will undoubtedly change following publication. It is the Reader's responsibility to obtain the most up-to-date information on any topic.

NAB and the contributors to this publication make no representations and disclaims all warranties, express or implied, (except as otherwise provided by law) regarding the information contained herein, including without limitation the implied warranty of fitness for a particular purpose and the implied warranty of merchantability. To the extent permitted by law, neither NAB nor any contributor to this publication shall be liable to any other person or entity for any damages, including without limitation any direct, indirect, special, consequential, exemplary or punitive damages under any theory of law.

ISBN 0-089324-258-6
Item # 3839

For more information on other publications available through the National Association of Broadcasters, call + 1 (202) 429-5373, (800) 368-5644 or check out NAB's website at http://www.nab.org.

FOREWORD

The 19th century writer/journalist Ambrose Bierce once dismissed an entire book in one nine-word review, stating: "The covers of this book are too far apart." This 9th edition of the *NAB Engineering Handbook* could perhaps also be so criticized, but only on the basis of the substantial shelf space it occupies. Inside, the material between the covers includes a tremendous amount of useful reference information, covering all aspects of broadcast engineering technology and practices. This Handbook is the definitive reference source on the subject of broadcast engineering, and well deserves and utilizes these widely displaced covers.

It has been six years since a new edition of the *NAB Engineering Handbook* was produced. In the current environment, complete product introduction cycles from research to production lines can span less than a year. In a perfect world, a new edition of the *NAB Engineering Handbook* might be produced on an annual basis, or even semi-annually, to attempt to truly keep pace with the rate of technology changes. But we don't live in a perfect world, do we? The compilation of this Handbook is a gargantuan task, requiring significant work effort from literally hundreds of people and thousands of hours of writing, editing and preparatory work. It is a multi-year effort of large proportion and it is no exaggeration to say that work on the 9th edition began not long after the 8th edition was published. All those who worked so tirelessly to bring this publication to life have our deepest appreciation and hearty congratulations.

The *NAB Engineering Handbook* is a snapshot in time of broadcast and broadcast-related technology. In that sense it is not perfect, because it is a snapshot, while the blistering pace of broadcast and telecommunications technology development and deployment continues. However, as our industry now finds itself in the initial stages of the digital broadcasting age, it is truly a perfect time to introduce this new edition of the *Handbook*. Many subjects related to the "digital revolution" that were not covered in the 8th edition, because they were at an early stage of development or only theorized about at that time, are dealt with in great detail in this edition.

NAB Science and Technology remains committed to being responsive to the dynamic, time-sensitive needs of the ever-evolving broadcast industry. Please feel free to let us know your views on this edition of the *Handbook* and your suggestions for future editions.

Lynn D. Claudy
Senior Vice President, Science and Technology
National Association of Broadcasters

ACKNOWLEDGEMENTS

"Its technical future is as limitless as space, its service potential is bounded only by the ingenuity of man, and its economic progress will keep step with that of the country itself." These words taken from the 5th edition of the *NAB Engineering Handbook* are as true today about the broadcasting industry as they were in 1960.

To prepare engineers for the technical future of broadcasting, we called upon the ingenuity of many to produce this 9th edition of the *NAB's Engineering Handbook*. First among those is Jerry Whitaker, the editor-in-chief of this edition. Jerry, who is the president of Technical Press and the former editor and publisher of *Broadcast Engineering Magazine,* has lent his talents to numerous technical broadcast publications. Jerry brought a truly fresh perspective and limitless enthusiasm to this publication.

It would be difficult for anyone to imagine that a publication like the *NAB Engineering Handbook* could be produced without the assistance of numerous individuals. We would like to express our appreciation to the members of the NAB Handbook Board of Advisors:

Andy Butler, PBS, Alexandria, VA
Randall Hoffner, ABC, New York, NY
Louis Libin, Broadcast Comm Inc., Woodmere, NY
Alfred Resnick, ABC, New York, NY
Milford Smith, Greater Media, East Brunswick, NJ
Donald Wilkinson, Fisher Broadcasting, Seattle, WA

Recognition is also due to the engineering staff of NAB's Science and Technology department for reviewing all of the manuscripts: John Marino, Art Allison, David Layer, David Wilson, and Kelly Williams.

Technical editing was handled by Courtenay Smith Brown. Text processing and author communications was led by Sharon Devine with Cheryl Coleridge providing additional text processing support. Finally, a special thanks to Lynn Claudy, Senior Vice President of NAB Science and Technology for supporting and encouraging this project, John Battison for lending his expertise to review manuscripts, Dos Schuette for contributing her design expertise, and Ann Carper for adding her editing skills to this publication.

Janet H. Elliott
Managing Editor

TABLE OF CONTENTS

SECTION 1. BROADCAST STANDARDS AND PRACTICES

SECTION 2. BROADCAST TOWERS AND SYSTEMS

NAB BROADCASTERS

SECTION 3. AUDIO PRODUCTION FACILITIES

SECTION 4. RADIO TRANSMISSION FACILITIES

SECTION 5. VIDEO PRODUCTION FACILITIES

SECTION 6. TELEVISION TRANSMISSION FACILITIES

SECTION 7. SIGNAL MEASUREMENT AND TESTING

SECTION 8. ANCILLARY BROADCAST SYSTEMS

SECTION 9. TECHNICAL MANAGEMENT AND SAFETY

Section 1: Broadcast Standards and Practices

Introduction by Al Resnick, ABC, New York, NY

By its very nature, broadcasting requires coordination and cooperation among stations and end-users. The fundamental basis of radio and television broadcasting is the application of mutually-agreeable standards and practices that permit stations to meet their charter—namely, serving the public interest, convenience and necessity. To accomplish this goal, the Federal Communications Commission (FCC) has established a body of standards and rules upon which broadcast services in the U.S. are based. Building upon this foundation, a number of organizations have developed additional policies that guide the broadcast industry, the most recent being the Advanced Television Systems Committee (ATSC), which built upon a rich history of standards work by the Society of Motion Picture and Television Engineers (SMPTE) and related organizations around the world to produce the new digital television (DTV) standard. It is the examination of these important issues that this section of the *NAB Engineering Handbook* is devoted.

John Norgard provides a comprehensive overview of the lifeline of the broadcast industry, in his chapter, *Electromagnetic Spectrum.*

William R. Meintel's *Frequency Allocations For Broadcasting and the Broadcast Auxiliary Services* chapter provides an overview of terrestrial frequency allocations and a listing of the frequencies available for AM, FM and TV broadcast stations as well as for the auxiliary broadcast services that support broadcasting operations.

If you are new to the broadcast industry, the Office of Public Affairs of the FCC contribution is an invaluable chapter. It is full of information on the Commission's organization and administrative practices.

In his chapter, *FCC Compliance and Information Bureau,* NAB staffer David E. Wilson explains the intricacies of the FCC bureau that is responsible for providing information to the general public, resolving telecommunications interference problems and, most importantly, enforcing the FCC's technical rules.

How broadcasters co-exist in the increasingly crowded spectrum is the focus of Robert Van Buhler's and Richard Rudman's *Frequency Coordination* chapter. Frequency coordination is necessary to facilitate the allocation of limited spectrum for a growing number of uses among a growing number of users.

Do you know the difference between the *flat-earth and spherical-earth* distance calculation methods? In his chapter, *Distance and Bearing Calculations,* Dane E. Ericksen provides an in-depth analysis of both methods.

NAB staffer Courtenay Smith Brown's *Broadcast Related Organizations and Information* chapter provides a general overview of entities whose activities affect the work of broadcast engineers.

As the broadcast industry enters the digital era, Chip Morgan and Randall Hoffner team up to provide the latest information on digital audio technology in their *Digital Audio Standards and Practices* chapter.

As we look forward to the digital age, it behooves us to consider our analog past. The definitive chapter on NTSC is presented here in a chapter titled *NTSC Standard.*

If film can be exchanged internationally, why can't we accomplish the same exchange with television programming? D.H. Pritchard and J.J. Gibson explore this lack of compatibility in their *Worldwide Television Standards for Conventional Television* chapter.

NAB staffers contribute to this opening section once again, in a chapter prepared by Arthur Allison and David H. Layer, entitled *Digital Television*

Terrestrial Transmission Standards. Their joint effort provides background information on the various aspects of DTV technology and introduces the reader to the standards that define DTV technology.

David E. Wilson's second offering to this section is the *EAS Standard* chapter. Wilson's chapter analyzes EAS from four different perspectives—the modulation method, the coding structure, how each party that participates in EAS relates to each of the other parties and the EAS responsibilities of individual broadcasters.

1.1
ELECTROMAGNETIC SPECTRUM

JOHN NORGARD
UNIVERSITY OF COLORADO, COLORADO SPRINGS, CO

INTRODUCTION

The electromagnetic (EM) spectrum consists of all forms of EM radiation [EM waves (radiant energy) propagating through space, from DC to light to gamma rays]. The EM spectrum can be arranged in order of frequency and/or wavelength into a number of regions,[1] usually wide in extent, within which the EM waves have some specified common characteristics, (*e.g.*, those characteristics relating to the production or detection of the radiation). A common example is the spectrum of the radiant energy in white light, as dispersed by a prism, to produce a "rainbow" of its constituent colors.

The EM spectrum is typically displayed as a function of frequency (or wavelength), as shown schematically in Figure 1.1-1. In air, frequency (f) and wavelength (λ) are inversely proportional ($f = c/\lambda$, where $c \approx 3 \times 10^8$ m/s is the speed of light in vacuum). The Meter Kilogram Second (MKS) unit of frequency is the Hertz (1 Hz = 1 cps); the MKS unit of wavelength is the meter. Frequency is also measured in the following sub-units:

Kilohertz, 1 kHz = 10^3 Hz
Megahertz, 1 MHz = 10^6 Hz
Gigahertz, 1 GHz = 10^9 Hz
Terahertz, 1 THz = 10^{12} Hz
Petahertz, 1 PHz = 10^{15} Hz
Exahertz, 1 EHz = 10^{18} Hz

or, for very high frequencies:

electron volts[2], 1 ev ~ 2.41×10^{14} Hz

Wavelength is also measured in the following sub-units:

Centimeters, 1 cm = 10^{-2} m
Millimeters, 1 mm = 10^{-3} m
Micrometers, 1 μm = 10^{-6} m (microns)
Nanometers, 1 nm = 10^{-9} m
Ångstroms, 1 Å = 10^{-10} m
Picometers, 1 pm = 10^{-12} m
Femtometers, 1 fm = 10^{-15} m
Attometers, 1 am = 10^{-18} m

SPECTRAL SUB-REGIONS

In this chapter, for convenience, the overall EM spectrum is divided into three main sub-regions:

Optical spectrum
DC to light spectrum
Light to gamma ray spectrum

This section focuses on the main sub-regions of the EM spectrum. Note that the boundaries between some of the spectral regions are somewhat arbitrary. Some spectral bands have no sharp edges and merge into each other and some spectral segments overlap each other slightly.

Optical Spectrum

The optical spectrum is the "middle" frequency/wavelength region of the EM spectrum. It is defined here as the visible and near-visible regions of the EM spectrum and includes:

- The infrared (IR) band, circa 300 μm–0.7 μm (circa 1 THz–429 THz)
- The visible light band, 0.7 μm–0.4 μm (429 THz–750 THz)
- The ultraviolet (UV) band, 0.4 μm–circa 10 nm (750 THz–circa 30 PHz, approximately 100 ev)

These regions of the EM spectrum are usually described in terms of their wavelengths.

Atomic and molecular radiation produce radiant light energy. Molecular radiation and radiation from hot bodies produce EM waves in the IR band. Atomic radiation (outer shell electrons) and radiation from arcs and sparks produce EM waves in the UV band.

Visible Light Band

In the "middle" of the optical spectrum is the visible light band, extending *approximately* from 0.4 μm (violet) up to 0.7 μm (red), (from 750 THz (violet)–429 THz (red)). EM radiation in this region of the EM spectrum, when entering the eye, gives rise to visual sensations (colors), according to the spectral response of the eye, which responds only to radiant energy in the visible light band extending from the extreme long wavelength edge of red to the extreme short wavelength edge of violet.[3] This visible light band is further subdivided into the various colors of the rainbow (in decreasing wavelength/increasing frequency):

- Red, a primary color, peak intensity at 700.0 nm (429 THz)
- orange
- yellow
- green, a primary color, peak intensity at 546.1 nm (549 THz)

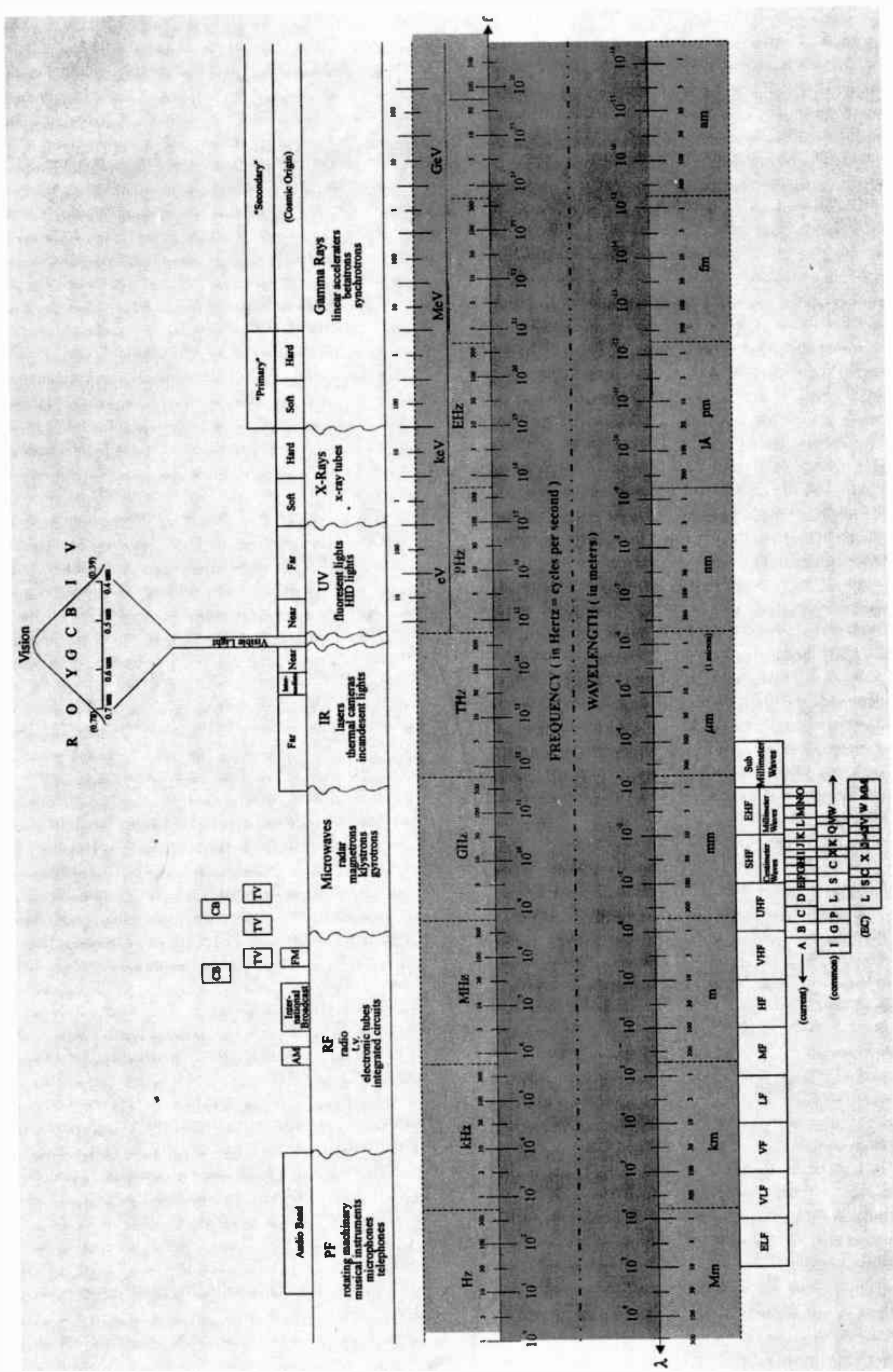


Figure 1.1-1. Simplified chart of the electromagnetic spectrum. (From, Whitaker, J. C., The Electronics Handbook, CRC Press, Boca Raton, FL, 1996. Used with permission.)

- cyan
- blue, a primary color, peak intensity at 435.8 nm (688 THz)
- indigo
- violet

IR Band

The IR band is the region of the EM spectrum lying immediately below the visible light band. The IR band consists of EM radiation with wavelengths extending between the longest visible red (circa 0.7 μm) and the shortest microwaves (300 μm–1 mm), (circa 429 THz to 1 THz–300 GHz).

The IR band is further subdivided into the "near" (shortwave), "intermediate" (midwave), and "far" (longwave) IR segments as follows: [4]

Near IR segment, 0.7 μm–3 μm (429 THz–100 THz)

Intermediate IR segment, 3 μm–7 μm (100 THz–42.9 THz)

Far IR segment, 7 μm–300 μm (42.9 THz–1 THz)

Note that the sub-millimeter region of wavelengths is sometimes included in the very far region of the IR band.

Sub-millimeter band, 100 μm up to 1 mm (3 THz–300 GHz)

EM radiation is produced by oscillating and rotating molecules and atoms. Therefore, all objects at temperatures above absolute zero emit EM radiation by virtue of their thermal motion (warmth) alone. Objects near room temperature emit most of their radiation in the IR band. However, even relatively cool objects emit some IR radiation; hot objects, such as incandescent filaments, emit strong IR radiation.

IR radiation is sometimes incorrectly called "radiant heat," because warm bodies emit IR radiation and bodies that absorb IR radiation are warmed. However, IR radiation is not itself "heat." This radiant energy is called "black body" radiation. Such waves are emitted by all material objects. For example, the background cosmic radiation (2.7K) emits microwaves, room temperature objects (293K) emit IR rays, the Sun (6000K) emits yellow light, the Solar Corona (1 Million K) emits X-rays.

IR astronomy uses the 1 μm to 1 mm part of the IR band to study celestial objects by their IR emissions. IR detectors are used in night vision systems, intruder alarm systems, weather forecasting and missile guidance systems. IR photography uses multilayered color film, with an IR sensitive emulsion in the wavelengths between 700–900 nm, for medical and forensic applications and for aerial surveying.

UV Band

The UV band is the region of the EM spectrum lying immediately above the visible light band. The UV band consists of EM radiation with wavelengths extending between the shortest visible violet (circa 0.4 μm) and the longest X-rays (circa 10 nm)[5], from 750 THz (approximately 3 ev) up to circa 30 PHz (approximately 100 ev).

The UV band is further subdivided into the "near" and the "far"[6] UV segments as follows:

Near UV segment, circa 0.4 μm down to 100 nm (circa 750 THz–3 PHz), approximately 3 ev up to 10 ev

Far UV segment, 100 nm down to circa 10 nm, (3 PHz–circa 30 PHz), approximately 10 ev up to 100 ev

UV radiation is produced by electron transitions in atoms and molecules, as in a mercury discharge lamp. Radiation in the UV range is easily detected and can cause florescence in some substances and can produce photographic and ionizing effects.

In UV astronomy, the emissions of celestial bodies in the wavelength band between 50–320 nm are detected and analyzed to study the heavens. The hottest stars emit most of their radiation in the UV band.

DC to Light

Below the IR band are the lower frequency (longer wavelength) regions of the EM spectrum, subdivided generally into the following spectral bands (by frequency/wavelength):

Microwave band, 300 GHz–300 MHz[7] (1 mm–1 m)

Radio frequency (RF) band, 300 MHz–10 kHz (1 m–30 Km)

Power(PF)/telephony band, 10 kHz–DC (30 Km–∞)

These regions of the EM spectrum are usually described in terms of their frequencies.

EM radiation whose wavelengths are of the order of millimeters and centimeters are called *microwaves*, and those still longer are called radio frequency (RF) waves (or *Hertzian waves*).

Radiation from electronic devices produces EM waves in both the microwave and RF bands. Power frequency energy is generated by rotating machinery. Direct current (dc) is produced by batteries or rectified alternating current (ac).

Microwave Band

The microwave band is the region of wavelengths between the far IR/sub-millimeter region and the conventional RF region. The boundaries of the microwave band have not been definitely fixed, but it is commonly regarded as the region of the EM spectrum extending from about 1 mm–1 m in wavelengths (300 GHz–300 MHz). The microwave band is further sub-divided into the following segments:

Millimeter waves,[8] 300 GHz–30 GHz (1 mm–1 cm), "EHF" band

Centimeter waves, 30 GHz–3 GHz (1 cm–10 cm), "SHF" band

The microwave band usually includes the "UHF" band from 3 GHz–300 MHz (10 cm–1 m). Microwaves are used in radar, in communication links spanning moderate distances, as radio carrier waves in radio broadcasting, for mechanical heating and cooking in microwave ovens.

Radio Frequency (RF) Band

The RF range of the EM spectrum is the wavelength band suitable for utilization in radio communications extending from 10 kHz–300 MHz,[9] (30 Km–1 m). Some of the radio waves serve as the carriers of low-frequency audio signals; other radio waves are modulated by video and digital information. The *amplitude modulated* (AM) broadcasting band uses waves with frequencies between 550–1640 kHz; the *frequency modulated* (FM) broadcasting band uses waves with frequencies between 88–108 MHz.

In the U.S., the Federal Communications Commission (FCC) is responsible for assigning a range of frequencies (*e.g.*, a frequency band in the RF spectrum) to a broadcasting station or service. The International Telecommunications Union (ITU) coordinates frequency band allocation and coordination on a worldwide basis.

Radio astronomy uses a radio telescope to receive and study radio waves naturally emitted by objects in space. Radio waves are emitted from hot gases ("thermal" radiation), from charged particles spiraling in magnetic fields (synchrotron radiation), and from excited atoms and molecules in space (spectral lines), such as the 21 cm line emitted by hydrogen gas.

Power Frequency (PF)/Telephone Band

The PF range of the EM spectrum is the wavelength band suitable for generating, transmitting, and consuming low frequency power, extending from 10 kHz down to DC (30 km to ∞). In the U.S., most power is generated at 60 Hz (some military and computer applications use 400 Hz); in Europe, power is generated at 50 Hz.

FREQUENCY BAND DESIGNATIONS

The combined microwave, RF and power/telephone spectra are subdivided into the specific bands shown below in Table 1.1-1. The radar band often is considered to extend from the middle of the HF (7) band to the end of the EHF (11) band. Note that the band designated (12) has no commonly used name or abbreviation.

The U.S. military Tri-Service designations for radio communications bands are as shown in Table 1.1-2.

Another set of designations that pre-date the U.S. Tri-Service designations are in Table 1.1-3. Sometimes, the prior S band extended from 1.5 GHz to 5.85 GHz, and the prior X band extended from 5.85 GHz to 10.9 GHz, eliminating the C band.

The "EC" radar band designations in prior use are contained in Table 1.1-4. The prior K band sometimes included the Ka band and extended from 18 to 40 GHz.

An alternate and more detailed sub-division of the UHF (9), SHF (10), and EHF (11) bands is shown in Table 1.1-5.

Table 1.1-1

International Radio Frequency Band Designations and the Numerical Designations

Band Name	Designation	Frequency Range	Wavelength Range
Extremely Low Frequency	ELF (1) Band	3 Hz–30 Hz	100,000 km–10,000 km
Super Low Frequency	SLF (2) Band	30 Hz–300 Hz	10,000 km–1,000 km
Ultra Low Frequency	ULF (3) Band	300 Hz–3 kHz	1,000 km–100 km
Very Low Frequency	VLF (4) Band	3 kHz–30 kHz	100 km–10 km
Low Frequency	LF (5) Band	30 kHz–300 kHz	10 km–1 km
Medium Frequency	MF (6) Band	300 kHz–3 MHz	1 km–100 m
High Frequency	HF (7) Band	3 MHz–30 MHz	100 m–10 m
Very High Frequency	VHF (8) Band	30 MHz–300 MHz	10 m–1 m
Ultra High Frequency	UHF (9) Band	300 MHz–3 GHz	1 m–10 cm
Super High Frequency	SHF (10) Band	3 GHz–30 GHz	10 cm–1 cm
Extremely High Frequency	EHF (11) Band	30 GHz–300 GHz	1 cm–1 mm
	(12) Band	300 GHz–3 THz	1 mm–100 μm

Table 1.1-2

Current U.S. Tri-Service Radar Band Designations

Band Designation	Frequency Range	Wavelength Range
A Band	0 Hz–250 MHz	(∞–1.2 m)
B Band	250 MHz–500 MHz	(1.2 m–60 cm)
C Band	500 MHz–1 GHz	(60 cm–30 cm)
D Band	1 GHz–2 GHz	(30 cm–15 cm)
E Band	2 GHz–3 GHz	(15 cm–10 cm)
F Band	3 GHz–4 GHz	(10 cm–7.5 cm)
G Band	4 GHz–6 GHz	(7.5 cm–5 cm)
H Band	6 GHz–8 GHz	(5 cm–3.75 cm)
I Band	8 GHz–10 GHz	(3.75 cm–3 cm)
J Band	10 GHz–20 GHz	(3 cm–1.5 cm)
K Band	20 GHz–40 GHz	(1.5 cm–7.5 mm)
L Band	40 GHz–60 GHz	(7.5 mm–5 mm)
M Band	60 GHz–100 GHz	(5 mm–3 mm)
N Band	100 GHz–200 GHz	(3 mm–1.5 mm)
O Band	200 GHz–300 GHz	(1.5 mm–1 mm)

Table 1.1-3

Band Designations that Pre-date the U.S. Tri-Service Designations

Band Designation	Frequency Range	Wavelength Range
I Band	100 MHz–150 MHz	(3 m–2 m)
G Band	150 MHz up to 225 MHz	(2 m–1.33 m)
P Band	225 MHz–390 MHz	(1.33 m–76.9 cm)
L Band	390 MHz–1.5 GHz	(7.69 cm–19.4 cm)
S Band	1.5 GHz–3.9 GHz	(19.4 cm–7.69 cm)
C Band	3.9 GHz–6.2 GHz	(7.69 cm–48.4 mm)
X Band	6.2 GHz–10.9 GHz	(48.4 mm–27.5 mm)
K Band	10.9 GHz–36 GHz	(27.5 m–8.33 mm)
Q Band	36 GHz–46 GHz	(8.33 mm–6.52 mm)
V Band	46 GHz–56 GHz	(6.52 mm–5.36 mm)
W Band	56 GHz – —	(5.36 mm — . . .)

Table 1.1-4
"EC" Radar Band Designations in Prior Use

Band Designation	Frequency Range	Wavelength Range
L Band	1 GHz–2 GHz	(30 cm–15 cm)
S Band	2 GHz–4 GHz	(15 cm–7.5 cm)
C Band	4 GHz–8.2 GHz	(7.5 cm–3.66 cm)
X Band	8.2 GHz–12.4 GHz	(3.66 cm–2.42 cm)
Ku Band	12.4 GHz–18 GHz	(2.42 cm–1.67 cm)
K Band	18 GHz–26.5 GHz	(1.67 cm–11.3 mm)
Ka Band	26.5 GHz–40 GHz	(11.3 cm–7.5 mm)
V Band	40 GHz–75 GHz	(7.5 mm–4 mm)
W Band	75 GHz–110 GHz	(4 mm–2.73 mm)
mm Band	110 GHz–300 GHz	(2.73 mm–1 mm)

Table 1.1-5
Alternate and More Detailed Sub-division of the UHF (9), SHF (10), and EHF (11) Bands

Band Designation	Frequency Range	Wavelength Range
L Band	1.12 GHz–1.7 GHz	(26.8 cm–17.6 cm)
LS Band	1.7 GHz–2.6 GHz	(17.6 cm–11.5 cm)
S Band	2.6 GHz–3.95 GHz	(11.5 cm–7.59 cm)
C(G) Band	3.95 GHz–5.85 GHz	(7.59 cm–5.13 cm)
XN(J, XC) Band	5.85 GHz–8.2 GHz	(5.13 cm–3.66 cm)
XB(H, BL) Band	7.05 GHz–10 GHz	(4.26 cm–3 cm)
X Band	8.2 GHz–12.4 GHz	(3.66 cm–2.42 cm)
Ku(P) Band	12.4 GHz–18 GHz	(2.42 cm–1.67 cm)
K Band	18 GHz–26.5 GHz	(1.67 cm–1.13 cm)
V(R, Ka) Band	26.5 GHz–40 GHz	(1.13 cm–7.5 mm)
Q(V) Band	33 GHz–50 GHz	(9.09 mm–6 mm)
M(W) Band	50 GHz–75 GHz	(6 mm–4 mm)
E(Y) Band	60 GHz–90 GHz	(5 mm–3.33 mm)
F(N) Band	90 GHz–140 GHz	(3.33 mm–2.14 mm)
G(A)	140 GHz–220 GHz	(2.14 mm–1.36 mm)
R Band	220 GHz–325 GHz	(1.36 mm–0.923 mm)

Table 1.1-6
Power Band

Band Designation	Frequency Range
Sub-sonic band	0 Hz–10 Hz
Audio band	10 Hz–10 kHz
Ultra-sonic band	10 kHz and up

Table 1.1-7
RF Band

Longwave broadcasting band	150-290 kHz
AM broadcasting band	550–1640 kHz (1.640 MHz), 107 channels, 10 kHz separation
International broadcasting band	3–30 MHz
Shortwave broadcasting band	5.95–26.1 MHz (8 bands)
VHF TV (Channels 2–4)	54–72 MHz
VHF TV (Channels 5–6)	76–88 MHz
FM broadcasting band	88–108 MHz
VHF TV (Channels 7–13)	174–216 MHz
UHF TV (Channels 14–69)	512–806 MHz

Table 1.1-8
Microwave Band (Up to 40 GHz)

Band Designation	Frequency Range
Aero Navigation	0.96–1.215 GHz
GPS Down Link	1.2276 GHz
Military COM/Radar	1.35–1.40 GHz
Miscellaneous COM/Radar	1.40–1.71 GHz
L-Band Telemetry	1.435–1.535 GHz
GPS Down Link	1.57542 GHz
Military COM (Troposcatter/Telemetry)	1.71–1.85 GHz
Commercial COM & Private LOS	1.85–2.20 GHz
Microwave Ovens	2.45 GHz
Commercial COM/Radar	2.45–2.69 GHz
Instructional TV	2.50–2.69 GHz
Military Radar (Airport Surveillance)	2.70–2.90 GHz
Maritime Navigation Radar	2.90–3.10 GHz
Miscellaneous Radars	2.90–3.70 GHz
Commercial C-Band SAT COM Down Link	3.70–4.20 GHz
Radar Altimeter	4.20–4.40 GHz
Military COM (Troposcatter)	4.40–4.99 GHz
Commercial Microwave Landing System	5.00–5.25 GHz
Miscellaneous Radars	5.25–5.925 GHz
C-Band Weather Radar	5.35–5.47 GHz
Commercial C-Band SAT COM Up Link	5.925–6.425 GHz
Commercial COM	6.425–7.125 GHz
Mobile TV Links	6.875–7.125 GHz
Military LOS COM	7.125–7.25 GHz
Military SAT COM Down Link	7.25–7.75 GHz
Military LOS COM	7.75–7.9 GHz
Military SAT COM Up Link	7.90–8.40 GHz
Miscellaneous Radars	8.50–10.55 GHz
Precision Approach Radar	9.00–9.20 GHz
X-Band Weather Radar (& Maritime Navigation Radar)	9.30–9.50 GHz
Police Radar	10.525 GHz
Commercial Mobile COM (LOS & ENG)	10.55–10.68 GHz
Common Carrier LOS COM	10.70–11.70 GHz
Commercial COM	10.70–13.25 GHz
Commercial Ku-Band SAT COM Down Link	11.70–12.20 GHz
DBS Down Link & Private LOS COM	12.20–12.70 GHz
ENG & LOS COM	12.75–13.25 GHz
Miscellaneous Radars & SAT COM	13.25–14.00 GHz
Commercial Ku-Band SAT COM Up Link	14.00–14.50 GHz
Military COM (LOS, Mobile, &Tactical)	14.50–15.35 GHz
Aero Navigation	15.40–15.70 GHz
Miscellaneous Radars	15.70–17.70 GHz
DBS Up Link	17.30–17.80 GHz
Common Carrier LOS COM	17.70–19.70 GHz
Commercial COM (SAT COM & LOS)	17.70–20.20 GHz
Private LOS COM	18.36–19.04 GHz
Military SAT COM	20.20–21.20 GHz
Miscellaneous COM	21.20–24.00 GHz
Police Radar	24.15 GHz
Navigation Radar	24.25–25.25 GHz
Military COM	25.25–27.50 GHz
Commercial COM	27.50–30.00 GHz
Military SAT COM	30.00–31.00 GHz
Commercial COM	31.00–31.20 GHz
Navigation Radar	31.80–33.40 GHz
Miscellaneous Radars	33.40–36.00 GHz
Military COM	36.00–38.60 GHz
Commercial COM	38.60–40.00 GHz

Several other frequency bands of interest (not exclusive) are shown in Tables 1.1-6, 1.1-7, and 1.1-8.

Light to Gamma Rays

Above the UV spectrum are the higher frequency (shorter wavelength) regions of the EM spectrum, subdivided generally into the following spectral bands (by frequency/wavelength):[10]

X-ray band, approximately 10 ev–1 Mev (circa 10 nm–circa 1 pm), circa 3 PHz–circa 300 EHz

Gamma ray band, approximately 1 Kev–∞ (circa 300 pm–0 m), circa 1 EHz–∞

These regions of the EM spectrum are usually described in terms of their photon energies in electron volts. Notice that the bottom of the gamma ray band overlaps the top of the X-ray band.

Radiation from atomic inner shell excitations produces EM waves in the X-ray band. Radiation from naturally radioactive nuclei produces EM waves in the gamma ray band.

X-Ray Band

The x-ray band is further sub-divided into the following segments:

Soft X-rays, approximately 10 ev–10 Kev (circa 10 nm–100 pm), circa 3 PHz–3 EHz
Hard X-rays, approximately 10 Kev–1 Mev (100 pm–circa 1 pm), 3 EHz–circa 300 EHz

Because the physical nature of these rays was first unknown, this radiation was called X-rays. The more powerful X-rays are called hard X-rays and are of high frequencies and, therefore, are more energetic; less powerful X-rays are called soft X-rays and have lower energies.

X-rays are produced by transitions of electrons in the inner levels of excited atoms or by rapid deceleration of charged particles (Brehmsstrahlung "breaking radiation"). An important source of X-rays is synchrotron radiation. X-rays can also be produced when high energy electrons from a heated filament cathode strike the surface of a target anode (usually tungsten) between which a high alternating voltage (approximately 100 kV) is applied.

X-rays are a highly penetrating form of EM radiation and applications of X-rays are based on their short wavelengths and their ability to easily pass through matter. X-rays are very useful in crystallography for determining crystalline structure and in medicine for photographing the body. Since different parts of the body absorb X-rays to a different extent, X-rays passing through the body provide a visual image of its interior structure when striking a photographic plate. X-rays are dangerous and can destroy living tissue. They can also cause severe skin burns. X-rays are useful in the diagnosis and non-destructive testing of products for defects.

Gamma Ray Band

The Gamma Ray band is sub-divided into the following segments:

Primary gamma rays, approximately 1 Kev–1 Mev (circa 300 pm–300 fm), circa 1 EHz–1000 EHz

The primary Gamma Rays are further sub-divided into the following segments:

Soft gamma rays, approximately 1 Kev–circa 300 Kev (circa 300 pm–circa 3 pm), circa 1 EHz–circa 100 EHz
Hard gamma rays, approximately 300 Kev–1 Mev (circa 3 pm–300 fm), circa 100 EHz–1000 EHz
Secondary gamma rays, approximately 1 Mev–∞ (300 fm–0 m), 1000 EHz–∞

Secondary gamma rays are created from collisions of high energy cosmic rays with particles in the Earth's upper atmosphere.

Gamma rays are essentially very energetic X-rays. The distinction between the two is based on their origin. X-rays are emitted during atomic processes involving energetic electrons; gamma rays are emitted by excited nuclei or other processes involving subatomic particles.

Gamma rays are emitted by the nucleus of radioactive material during the process of natural radioactive decay as a result of transitions from high energy excited states to low energy states in atomic nuclei. Cobalt 90 is a common gamma ray source (with a half-life of 5.26 years). Gamma rays are also produced by the interaction of high energy electrons with matter. "Cosmic" gamma rays cannot penetrate the Earth"s atmosphere.

Applications of gamma rays are found both in medicine and in industry. In medicine, gamma rays are used for cancer treatment, diagnoses, and prevention. Gamma ray emitting radioisotopes are used as tracers. In industry, gamma rays are used in the inspection of castings, seams and welds.

Defining Key Terms

Cosmic "rays"—Highly penetrating particle "rays" from outer space. Primary cosmic "rays" which enter the Earth's upper atmosphere consist of mainly protons. Cosmic "rays" of low energy have their origin in the sun, those of high energy in galactic or extragalactic space, possibly as a result of supernova explosions. Collisions with atmospheric particles result in secondary cosmic "rays" (particles) and secondary gamma rays (EM waves).

Electromagnetic spectrum—EM radiant energy arranged in order of frequency and/or wavelength and divided into regions within which the waves have some common specified characteristics (*e.g.*, the waves are generated, received, detected, or recorded in a similar way).

Gamma rays—Electromagnetic radiation of very high energy (> 30 Kev) emitted after nuclear reactions or by a radioactive atom when its nucleus is left in an excited state after emission of alpha or beta particles.

Infrared (IR) radiation—Electromagnetic radiation whose wavelengths lie in the range from 0.7 nm (the long-wavelength limit of visible red light) to 1 mm (the shortest microwaves). A convenient sub-division is as follows: Near: 0.7 μm to 2–5 μm; intermediate: 2–5 μm to 10 μm; far: 10 μm to 1 mm.

Light—White light, when split into a spectrum of colors, is composed of a continuous range of merging colors: red, orange, yellow, green, cyan, blue, indigo, violet.

Microwaves—An electromagnetic wave which has a wavelength between approximately 0.3 cm (or 1 mm) and 30 (or 10) cm, corresponding to frequencies between 1 GHz (or 300 MHz) and 100 (or 300) GHz. Note that there are no well defined boundaries distinguishing microwaves from infrared and radio waves.

Radio waves—Electromagnetic radiation suitable for radio transmission in the range of frequencies from about 10 kHz to about 300 MHz.

Ultraviolet (UV) radiation—Electromagnetic radiation whose wavelengths lie in the range from 0.4 nm (the shortest-wavelength limit of visible violet light) to 3 nm (the longest X-rays). A convenient sub-division is as follows: Near: 0.4 μm to 100 nm; far: 100 nm to 3 nm.

X-rays—Electromagnetic radiation of short wavelengths (circa 3 nm to 30 pm) produced when cathode rays impinge on matter.

Abbreviations

COM—Communications
DBS—Direct Broadcast Satellite
ENG—Electronic News Gathering
GPS—Global Positioning System
LOS—Line of Sight
SAT—Satellite

BIBLIOGRAPHY

Collocott, T. C., Dobson A. B., and Chambers W. R., *Dictionary of Science & Technology Eds. Handbook of Physics*, McGraw-Hill, New York, 1958.

Judd, D. B. and Wyszecki G., *Color in Business, Science and Industry*, 3rd Edition, John Wiley & Sons, New York.

Kaufman, Ed., *IES Illumination Handbook*, Illumination Engineering Society.

Lapedes, D. N., Editor, McGraw-Hill, *The McGraw-Hill Encyclopedia of Science & Technology*, 2nd Edition, New York.

Stemson, A., *Photometry and Radiometry for Engineers*, John Wiley & Sons, New York.

The Cambridge Encyclopedia, Cambridge University Press, 1990.

The Columbia Encyclopedia, Columbia University Press, 1993.

Webster's New World Encyclopedia, Prentice Hall, 1992.

Wyszecki G., and Stiles, W. S., *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd Edition, John Wiley & Sons, New York.

REFERENCES

1. Note that specific frequency ranges are often called "bands;" several contiguous frequency bands are usually called "spectrums;" and sub-frequency ranges within a band are sometimes called "segments."
2. Note that $\varepsilon = hf = eV$ where $h = 6.63(10^{-34}$ Js (Planck's Constant) and $e = 1.602 \times 10^{-19}$ C.
3. The spectral response of the eye is sometimes described as extending from 0.38 μm (violet) up to 0.75 or 0.78 μm (red), i.e. from 789 THz down to 400 or 385 THz).
4. Some reference texts use 2.5 μm (120 THz) as the breakpoint between the near and the intermediate IR bands and 10 μm (30 THz) as the breakpoint between the intermediate and the far IR bands. Also, 15 μm (20 Thz) is sometimes considered as the long wavelength end of the far IR band.
5. Some reference texts use 4, 5, or 6 nm as the upper edge of the UV band.
6. The far UV band is also referred to as the *vacuum UV band*, since air is opaque to all UV radiation in this region.
7. Some reference works define the lower edge of the microwave spectrum at 1 GHz.
8. Some reference articles consider the top edge of the millimeter region to stop at 100 GHz.
9. Some authors consider the RF band as extending from 10 kHz to 300 GHz, with the microwave band as a subset of the RF band from 300 MHz to 300 GHz.
10. Note that *cosmic "rays"* (from astronomical sources) are not EM waves (rays) and, therefore, are not part of the EM spectrum. Cosmic "rays" are high energy charged particles (electrons, protons, and ions) of extraterrestrial origin moving through space, which may have energies as high as 10^{20} ev. Cosmic "rays" have been traced to cataclysmic astrophysical/cosmological events, such as exploding stars and black holes. Cosmic "rays" are emitted by supernova remnants, pulsars, quasars, and radio galaxies. Cosmic "rays" which collide with molecules in the Earth's upper atmosphere produce secondary cosmic "rays" and gamma rays of high energy which also contribute to the natural background radiation. These gamma rays are sometimes called "cosmic" or *secondary* gamma rays. Cosmic rays are a useful source of high-energy particles for experiments.

1.2
FREQUENCY ALLOCATIONS FOR BROADCASTING AND THE BROADCAST AUXILIARY SERVICES

WILLIAM R. MEINTEL
TECHWARE, INC., CHANTILLY, VIRGINIA

INTRODUCTION

This chapter provides an overview of terrestrial frequency allocations and a listing of the frequencies available for AM, FM, and TV broadcast stations as well as for the "auxiliary" broadcast services that support broadcasting operations.[1] Necessarily, such an overview needs to include a description of the decision-making process that is involved in allocating a frequency band for a specific purpose.

This complex subject of allocations involves more than just the location of the service in the frequency spectrum; it includes decisions as to the number and width of channels, power and antenna limitations, as well as decisions concerning the technical standards that define how the spectrum will be shared with other users. Because each broadcast service presents unique service and interference objectives, the allocation process has been and will continue to be different for each of these services.

Here it is possible to provide only a brief description of the allocation process before turning to the current situation for each of the services. Where appropriate, the discussion will touch on changes expected in the foreseeable future.

HOW SPECTRUM IS ALLOCATED

It is a fundamental characteristic of radio wave propagation that these waves follow the laws of physics and thus ignore political or geographic boundaries. As a result, decisions concerning radio frequency allocations cannot be made solely at the local level but must take into account their anticipated impact outside the station's coverage area. Recognizing that coordination in the allocation and use of spectrum is essential, an international mechanism has been established to perform this function, and in the United States, a parallel coordination system has been established at the federal level.

[1] Technically the term "allocation" refers to the process by which a frequency band is made available for a specific purpose. However, it is used here in a broader sense that includes the "allotment" of frequencies within a band, and the "assignment" of individual stations.

International Allocation

At the international level, frequency allocation decisions are made by the International Telecommunication Union (ITU), a specialized agency of the United Nations, headquartered in Geneva, Switzerland. Like the United Nations, the ITU is a consortium of more than 180 governments, whose purposes are to propose, develop, revise and administer world-wide frequency allocation plans. In the terminology of the ITU, participating governments are referred to as "states," "members" or "administrations." Such international cooperation serves to minimize interference and maximize use of the spectrum.

Although the ITU maintains a permanent staff, the power of the organization, as indicated by the structure shown in Figure 1.2-1, resides with the member nations who direct its activities through periodic meetings. At the highest level of this structure is the Plenipotentiary Conference that adopts the fundamental policies of the organization and decides on the organization and activities of the Union through a treaty known as the *International Telecommunication Constitution and Convention*. Plenipotentiary Conferences are held at four year intervals and are open to ITU member countries, the United Nations and its specialized agencies, the International Atomic Energy Agency, regional telecommunications organizations and intergovernmental satellite operators.

Proceeding down the structure, the ITU Council consists of a representative group of Members elected by the Plenipotentiary Conference. The task of the Council is to insure that in the period between Plenipotentiary conferences, that the ITU's policies and strategy are adequately responding to the constantly changing telecommunication environment. The Council is also responsible for insuring efficient coordination of work, effective financial management and assisting Members in the implementation of the provisions and regulations of the Union.

The day-to-day work of the ITU is performed in the three sectors of which the Radiocommunications Sector is relevant to this discussion. The other sectors, as their names imply, deal with standardization and development of telecommunications.

World-wide allocations for radio services are made through decisions made at international conferences of the Radiocommunications Sector called World Radio

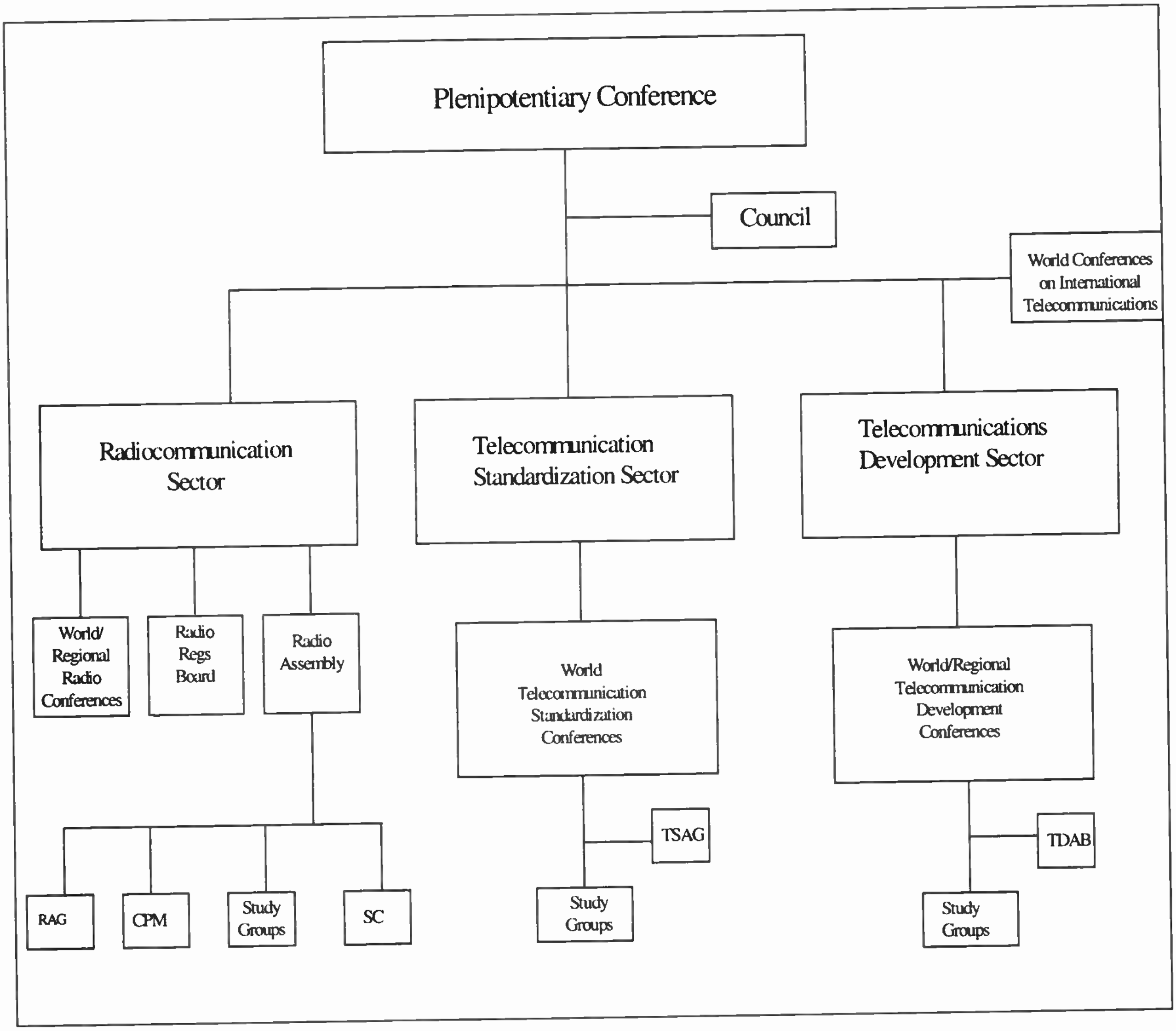


Figure 1.2-1. Organizational structure of the ITU.

Conferences (WRCs). These WRCs deal with frequency allocations and related matters of concern on a world-wide basis through review and revision of the international Radio Regulations. Through a process of give and take, decisions, are made for the allocation of spectrum. WRCs are held at two year intervals and deal with a limited agenda that is generally established four years in advance with a final agenda established by the ITU Council two years prior to the WRC. The early establishment of the agenda is intended to allow for proper preparation so that the work of the conference can be concluded in a timely manner; however, due to the rapid changes occurring in telecommunications, last minute requests for agenda changes are not uncommon.

Because of the geographic separation of certain parts of the world it was determined to be practical to divide the world into three regions. Region 1 consists of Europe and Africa; Region 2 consists of North and South America, Greenland and the Caribbean, and Region 3 consists of Asia and Oceania. In view of this, the ITU sometimes convenes Regional Radio Conferences (RRCs) to consider questions that are unique to a specific ITU Region. Often these RRCs consider implementation of decisions made at an earlier WRC.

Technical Submissions

Radiocommunications Assemblies are held in conjunction with and prior to the World Radio Conferences and provide the technical basis for the work of the WRC.[2] The Assemblies also approve and set

[2] In a reorganization of the ITU, the work of developing recommendations and providing reports dealing with technical issues formerly performed by the International Radio Consultative Committee (CCIR), that some readers may be familiar with, is now performed by the Radiocommunications Assemblies.

priorities for the work of the study groups that operate under the Assemblies' supervision. In addition, the Assemblies also setup or abolish study groups as deemed appropriate. This work is carried out by eight study groups, each of which is devoted to a particular radio communication service or specific technical issue.[3]

The work of two study groups is relevant to Broadcasting: Study Group 10 (Broadcasting Service—Sound) and study group 11 (Broadcasting Service—Television).[4] Under both of these Study Groups, there are sub-groups to deal with specific areas of interest. There also are joint study groups that deal with questions that are of relevance to more than one study group. Each study group has an international chairperson and one or more vice-chairmen who are provided by interested participating administrations.

ITU study group work is carried out by the members of the ITU as well as recognized users and standard setting groups such as the European Broadcasting Union and the U.S. broadcasting networks. Within individual administrations, organized structures often are created to provide input for the administration to submit to the Radio Assembly through its Study Groups. In the United States, this activity is chartered by and operates under the Department of State (DOS). There is a U.S. National Chairperson appointed by DOS, and two National Vice Chairpersons one from the Federal Communications Commission (FCC) and one from the National Telecommunications and Information Administration (NTIA). The DOS also appoints a U.S. National Chairperson for each of the individual study groups. Each is responsible for the work of the group and heads the U.S. delegation to international meetings of the study group. In the United States, study group activities are open to participation by the public. Contributions to study group work in the United States come mainly from the private sector. Elsewhere in the world, such work is performed primarily by government employees.

In general, the work of the study groups is timed to coincide with scheduled WRCs and RRCs with an international meeting of each study group usually occurring about once a year. In the interim, work is carried out by correspondence. Administrations may suggest any matter of interest for study that is within the purview of the Radiocommunications Sector, but priority is given to issues that are relevant to a scheduled WRC or RRC.

The cumulative recommendations and reports of the various study groups are updated and published periodically in hard copy books commonly referred to as the "Green Books." However, in recent years the ITU has begun providing material in electronic format and in a more timely manner by making individual recommendations available as they are approved. Interested parties may obtain information on obtaining these as well as numerous other ITU publications, in both hard copy and electronic form, from the ITU website (http://www.itu.int). The study group reports and recommendations contain a wealth of information relating not only to spectrum allocation but also related information concerning measurement procedures and standards for audio and video recording equipment used for the exchange of broadcast programming.

The Allocation Process

The allocation of radio frequency spectrum occurs as a result of a series of interrelated decisions. On the first, most basic level, blocks of frequencies are allocated on a worldwide basis by the ITU WRC process. Exactly which blocks of frequencies are allocated to particular services is determined by evaluating the many specific proposals submitted to the WRC for each frequency band. Technical input is obtained from propagation studies and other engineering analyses undertaken as part of the study group process and from the submissions of individual administrations. Frequently, blocks of spectrum are allocated for the same purpose on a worldwide basis, but may also be allocated for different purposes on a regional basis.

Member nations theoretically retain the sovereign right to domestic use of the spectrum so long as such use is not in contravention of the international Radio Regulations or the international agreements to which that administration is a party. However, as a practical matter, the flexibility of administrations to use the spectrum is limited by the worldwide allocation system and the need to avoid harmful interference.

Unlike the FCC, the ITU does not license users of the spectrum. Instead, it operates only as a coordinator, maintaining a Master International Frequency Register (MIFR or Master Register) of radio stations worldwide maintained by the Radiocommunications Bureau within the Radiocommunications Sector. Member administrations have agreed to provide notifications of new stations or modifications in existing stations operating within their respective countries. The Radiocommunications Bureau studies these notifications for compliance with the existing world or regional agreements and provides the results of its studies to the member nations. Only those notifications that comply with the existing agreements are placed in the Master Register of stations. Once a station has been placed in the Master Register, the member nations are obliged to provide it with the internationally agreed level of interference protection.

Because the WRC and RRC agreement texts provide only a general framework, many specific matters are left to individual nations to resolve and implement. In the U.S., the FCC and NTIA share the responsibility for implementing agreements to which the U.S. has assented, including the bilateral or multilateral agreements negotiated with our neighbors to deal with concerns that are unique to the countries involved.

[3] A complete list of the Radiocommunications Study Groups is contained in Appendix A.

[4] The numbering of the study groups is a carry over from the old CCIR structure.

There are differences in the treatment of the various broadcast services. Because lower frequency signals, such as those used by AM broadcasting, propagate over great distances, international decisions have a much greater impact on AM broadcasting than FM and television, whose VHF or UHF propagation is much more limited. This means that the restrictions on FM and TV allocations imposed by international agreements are usually applied only to areas near the borders. However, international agreement on technical transmission standards often is desired in order to foster the absence of interference and the world-wide free flow of communication.

Domestic Allocation Process

Regulation of spectrum began with the U.S. Department of Commerce in the early 1920s, when the Secretary of Commerce granted the first AM broadcasting licenses. By 1927, the number of AM stations had increased to 733, and over 6,000,000 radio receivers had been manufactured. However, because of an unfortunate court decision that precluded the Secretary of Commerce from dealing with the specific choice of location, power, and operating frequency, these matters were left largely to the discretion of the broadcaster. This led to a chaotic use of the spectrum, with widespread interference, a situation that led to the creation of the Federal Radio Commission in 1927. Seven years later the Federal Radio Commission was replaced by the FCC, formed pursuant to the Communications Act of 1934. Ever since, anyone desiring to operate a broadcast station, or most any kind of radio transmitting device, must apply to the FCC and be granted a license before commencing operation. Today, a broadcast license sets forth all essential technical parameters of station operation. The NTIA performs a similar function and coordinates the spectrum used by government agencies. The FCC works with the NTIA where there is a need for coordination between government and private uses of the spectrum.

Spectrum for use domestically, must be allocated by FCC rule making proceedings. Domestic allocation rule making proceedings can be initiated by the FCC on its own motion or in response to requests from the public, but in so doing it must not act in contravention of international agreements to which the United States is a party. Rule making proceedings are based on a public record developed through responses to the issuance of a *Notice of Proposed Rule Making* that are filed by interested parties. In addition to filing comments on the FCC's proposal, the public may reply to the comments of other parties. After the Commission evaluates the responses to its *Notice*, it may decide to either adopt the proposal as originally set forth, or modify it, based on the comments received, or possibly reject the proposal.[5]

Many FCC proceedings are controversial in nature. Allocations proceedings may be especially controversial, since a particular communications industry's livelihood may depend in part on how much spectrum is allocated. Thus, FCC allocation decision making is not simply a matter of technical evaluation but must be seen as part of the political process as well, as happens when entire industries may compete for a limited amount of spectrum. Where the number of users in a particular frequency band is expected to be relatively small or their use is sporadic, the FCC may propose sharing of this spectrum with other users. Such proposals may also be controversial, since sharing spectrum with a dissimilar service invites the possibility of interference and difficulties in coordinating the use of the frequencies. Because of these and other factors, FCC allocations proceedings consume a great deal of its time and energy and can impose burdens on the organizations that participate in them.

From time to time, alternative methods have been proposed for allocating spectrum. These ideas usually envision the removal of the FCC as the arbiter of mutually exclusive requests for spectrum and, instead substituting marketplace forces. Under a market allocation system, frequencies would become used by entities who would pay the most for them; social, or public policy aspects would not be considered. In recent years such a system has been employed for some frequency allocations in the form of spectrum auctions. In these cases the Commission's role is reduced to that of a technical "traffic cop" of the airwaves.

UNITED STATES AM BROADCASTING FREQUENCY ALLOCATION

In the United States, amplitude-modulated (AM) stations operate with carrier frequencies spaced 10 kHz apart. For many years the AM broadcast band in the United States included 107 channels, spaced 10 kHz apart, in the band from 535–1605 kHz. However, the frequency band allocated to AM broadcasting now includes a total of 117 channels in the band 535–1705 kHz, as shown in Figure 1.2-2.

Currently, there are over 4,700 commercial and noncommercial AM stations operating in the United States. These stations operate with various power levels, up to a maximum of 50 kW. About half of these stations use multi-tower directional antennas to restrict radiation in certain directions, for the purpose of controlling interference or maximizing radio service in particular directions.

Allocation decisions for the AM broadcast band are probably the most complex of the broadcast services. Since propagation varies with time of day, geographic latitude, soil conductivity and frequency, the engi-

535 kHz	-	1705 kHz
AM BROADCASTING		

Figure 1.2-2. The limits of the AM broadcast band.

[5] The nature of the Commission's rule making process is described in greater detail in Chapter 1.3.

neering analyses necessary to establish interference protection for other stations can be quite complicated. Engineers, the FCC and the ITU have sophisticated computer programs that analyze the input of a new or modified AM station proposal. Before going into the details of the AM broadcast allotment system currently in place in the United States, it is necessary to provide a brief history of AM broadcasting allocations.

Over the years, there have been many changes in the nature of AM broadcasting in the United States. In early years, clear channel stations provided the only service available in many areas of the U.S., but with the end of World War II, demand increased greatly and many AM stations were established in all areas of the country. Because of its early development of AM radio service, the U.S. experience has been used as a model for regional and bilateral agreements.

Early in the history of AM broadcasting, the countries in the North American area recognized the need to cooperate in the use of AM frequencies, and in 1937 they reached agreement on how to proceed. Soon, however, this agreement was found to be inadequate and negotiations began on a new agreement. Although the North American Regional Broadcasting Agreement (NARBA) was signed on November 15, 1950, it did not go into effect until 10 years later, on April 16, 1960. Signatories to NARBA include the United States, Canada, Cuba, the Dominican Republic and the United Kingdom on behalf of Jamaica and the Bahama Islands. Mexico, an earlier participant, removed itself from these negotiations, and a bilateral agreement between the U.S. and Mexico was reached in 1957.

These international agreements became necessary principally because nighttime AM propagation has the potential for causing widespread interference to neighboring countries unless mutual allocations criteria and related technical standards could be agreed upon and implemented by the parties to the agreement. To this end, NARBA provided for a partitioning of AM broadcast channels into three basic classes. The first of these are the so called "clear" channels, whose high-powered stations would have primary access to the frequency and other stations could use the channel subject to providing full protection to the dominant station(s). Clear channel dominant stations were designed to provide service over extensive areas by means of skywave as well as ground wave signals. NARBA set aside 60 of the 107 channels then available for clear channel use. Each NARBA country, except Jamaica, received a "priority" on one or more clear channels, with the U.S. receiving a major portion of available priorities. NARBA countries without a priority on a given clear channel could still assign stations on that channel, provided that these stations protected the wide-area service of the dominant station in the country with the NARBA priority.

The second class of channels were called "regional" channels and these occupied an additional 41 channels. Unlike clear channels, these channels were shared on an equal basis by all the NARBA countries. Stations operating on these channels were intended to provide service to a considerable area, but, unlike the clear channel stations that received protection for their skywave as well as groundwave service, only the groundwave service provided by the "regional" stations was protected.

The remaining six frequencies were the "local" channels that provide an even more limited type of groundwave service. Only limited interference protection was provided to these relatively low power operations.

Recognizing the need for updating these agreements and for developing more efficient coordination throughout Region 2 (North and South America), the 1979 General WRC called for a conference to be held in Region 2 to address AM broadcasting and sharing criteria. That conference was held in two sessions in 1980 and 1981 and resulted in the adoption of an agreement among most of the countries of the hemisphere. Included as part of that agreement, referred to as the 1981 Rio Agreement (for Rio de Janeiro, the location of the second conference session), was a list of all of the operating stations in the hemisphere along with information indicating whether or not the stations were receiving or causing harmful interference according to the technical criteria set forth in the agreement. Stations not causing interference were placed in the ITU Master Register and accorded protection from interference as defined by the agreement. In situations where interference already existed, the countries involved were asked to meet and work out mutually satisfactory solutions.

Because the general framework of the Rio Agreement did not deal with the particular needs and desires of the United States, Canada and Mexico, separate new bilateral agreements have been negotiated which incorporate the required additional items concerning coordination and technical parameters. Although the Rio Agreement applies throughout most of Region 2, as of this writing, relations with the Bahamas and the Dominican Republic continue to be governed by NARBA, as neither country has taken the necessary steps to replace the NARBA provisions with the Rio 1981 Agreement. Relations with Cuba regarding the 535–1605 kHz band are governed solely by the international Radio Regulations rather than by agreement.

The AM broadcasting system in the United States has continued to evolve as the demand for more stations has grown. One result of the demand for facilities is the increased use of directional antennas to provide required interference protection while enhancing coverage in other directions. With the spectrum becoming crowded, another development was a large increase in the number of daytime-only stations. These are stations that were authorized to operate only during daylight hours. Since propagation conditions during these daylight hours do not normally support significant skywave transmission, there are many locations where a station can be operated during the daytime without causing harmful interference to other stations. Based on this concept, the FCC over the years licensed approximately 1200 stations for daytime-only operation.

Recognizing that daytime-only stations (and even some full-time stations that operate with restrictive directional antenna patterns during the night) are unable to provide effective service during early morning hours, the FCC originally allowed these stations to operate during this period so long as no interference complaint had been received. Ultimately, this proved to be unworkable, and a more formalized approach was adopted. The FCC began granting Pre-Sunrise Authorizations (PSRAs), permitting many of these stations to operate with their daytime facilities with powers up to 500 W during the pre-sunrise period between 6:00 AM and local sunrise. While some interference occurred, the FCC believed the interference was balanced by the public's need for local informational services during this very important morning time period.

While pre-sunrise operation did provide some relief, it did not end the economic problems many stations faced in effectively competing with full-time stations. The FCC was pressed to provide relief in the form of post-sunset operation for daytime-only stations. With the removal of international impediments, the FCC did provide such relief. Once again, taking into consideration the need for more service of a local nature, and recognizing the changing nature of propagation conditions in which full nighttime conditions do not exist until several hours after sunset, the FCC granted Post Sunset Authorizations (PSSA) for most of the daytime only stations. These authorizations permit operation for periods of up to two hours past sunset with power reduced to prevent interference. The FCC also changed its rules concerning the minimum power at which a station is permitted to operate. This, in turn, led to a subsequent decision allowing many PSSA stations to operate throughout the night, albeit with reduced power.[6]

Still another change that has occurred pertained to the use of the Class I-A clear channels. At one time, only a single station was permitted to operate on these channels at night, but in two FCC decisions, first some and now all of these channels have been "broken down" to permit the authorization of additional nighttime operations. Clear channel stations that at one time provided service for a major portion of the country during nighttime hours now are protected only out to a distance averaging 750 miles.

Use of the Expanded Band

At an ITU Regional Radio Conference that concluded in 1988, participating administrations reached an agreement on the criteria for expanding the AM band in Region 2 by adding 10 new channels between 1605 to 1705 kHz. Although the regional agreement established specific technical criteria for the implementation of the new channels, (including the granting of priority usage of certain channels) the United States still retains considerable latitude in its domestic implementation. In a large country like the U.S., the use of the channels is unrestricted except in the relatively few areas near the borders with our neighbors.

The basic criteria set forth in the Regional Agreement are as follows: stations may operate with 1 kW of power with a non-directional antenna height of 90°, or stations may operate with a power not in excess of 10 kW by employing a directional antenna to provide equivalent protection to stations in other countries. The channels allotted to the United States in the border areas vary from location to location, but, as noted, over a large portion of the U.S. all ten of the channels may be used. Non-allotted channels are not precluded from use, but the allotted channels in the other countries must remain fully protected.

After a lengthy proceeding, the FCC developed a system whereby the channels in the expanded band would be used to improve the current interference situation by giving selected stations the opportunity to move to the new band. In view of this, applications for new stations are not currently being accepted for operation in the band 1605–1705 kHz.

Persons seeking an authorization for a new AM broadcast station on frequencies below 1605 kHz may do so by filing an application with the FCC. Applications must provide documentation that the proposed operation will comply with all applicable FCC rules as well as the appropriate international regulations. The details and methodology for allocating AM radio stations along with basic design specifications for AM directional antennas can be found in FCC Rules at Sections 73.14 to 73.190.

The 1981 Rio Agreement changed the NARBA station classifications. No longer are the channels themselves classified. Stations are now classified without regard to the channel on which they operate. Stations providing wide area service, both groundwave and skywave, are now designated Class A stations, while stations providing the equivalent of the regional and local services are designated as Class B and C respectively. The new agreement permits any class of station to operate on any channel so long as it provides protection to other stations based on their classification. Figures 1.2-3, 1.2-4 and 1.2-5 show the relationship

International classes of M stations	Corresponding U.S. classes prior to 1990	Classes of Channels available in U.S. for each class
Class A	I-A	Clear Channels
	I-B	Clear Channels
	I-N	Clear Channels
Class B	II	Clear Channels
	II-A	Clear Channels
	II-B	Clear Channels
	II-C	Clear Channels
	II-D	Clear Channels
	II-S	Clear Channels
	III	Regional channels
	III-S	Regional channels
Class C	IV	Local channels

Figure 1.2-3. International and domestic classifications of stations and channels prior to 1990.

[6] The rules pertaining to PSSA and PSRA are contained in FCC Rules and Regulations Section 73.99.

Frequency	Class of Station*	Frequency	Class of Station*	Frequency	Class of Station*
640 kHz	I-A	650 kHz	I-A	660 kHz	I-A
670 kHz	I-A	680 kHz	I-B	700 kHz	I-A
710 kHz	I-B	720 kHz	I-A	750 kHz	I-A
760 kHz	I-A	770 kHz	I-A	780 kHz	I-A
810 kHz	I-B	820 kHz	I-A	830 kHz	I-A
840 kHz	I-A	850 kHz	I-B	870 kHz	I-A
880 kHz	I-A	890 kHz	I-A	940 kHz	II
000 kHz	I-B	1020 kHz	I-A	1030 kHz	I-A
1040 kHz	I-A	1060 kHz	I-B	1070 kHz	I-B
1080 kHz	I-B	1090 kHz	I-B	1100 kHz	I-A
1110 kHz	I-B	1120 kHz	I-A	1130 kHz	I-B
1140 kHz	I-B	1160 kHz	I-A	1170 kHz	I-B
1180 kHz	I-A	1190 kHz	I-B	1200 kHz	I-A
1210 kHz	I-A	1500 kHz	I-B	1510 kHz	I-B
1520 kHz	I-B	1530 kHz	I-B	1540 kHz	I-B
1550 kHz	II	1560 kHz	I-B		

* In addition to the Class I-A or Class I-B stations that could be assigned to the above channels, various Class II stations could also be assigned.

REGIONAL CHANNELS

(Class III or III-S stations were assigned to these channels)
550, 560, 570, 580, 590, 600, 610, 620, 630, 790, 910, 920, 930, 950, 960, 970, 980, 1150, 1250, 1260, 1270, 1280, 1290, 1300, 1310, 1320, 1330, 1350, 1360, 1370, 1380, 1390, 1410, 1420, 1430, 1440, 1460, 1470, 1480, 1590, and 1600 kHz

LOCAL CHANNELS

(Class IV stations were assigned to these channels)
1230, 1240, 1340, 1400, 1450, and 1490 kHz

Figure 1.2-4. Channel utilization in the United States prior to 1990 clear channels.

between these international classifications and those that existed in the United States prior to 1990. In 1990, the FCC proposed to align the U.S. domestic classification system for AM stations with the 1981 Rio Agreement system. In addition to the three international classes of stations the FCC has added a fourth, Class D. Class D stations are those stations that operate nighttime with a power of less than 250 W and an equivalent RMS antenna field of less than 141 mV/m at 1 km. Such stations are usually former daytime only stations that have been granted some limited amount of nighttime operation. The current status of AM station classification is contained in Section 73.21 of the FCC Rules and Regulations. In that the allocation system

Class of Station	Power in kW Daytime hours Minimum	Maximum	Power in kW Nighttime hours Minimum	Maximum
IA	10.0	50.0	10.0	50.0
I-B	10.0	50.0	10.0	50.0
I-N	10.0	50.0	10.0	50.0
II	10.0	50.0	10.0	50.0
II-A	10.0	50.0	10.0	50.0
II-B	0.25	50.0	0.25	50.0
II-C	0.25	50.0	0.25	1.0
II-D	0.25	50.0	*	*
II-S	0.25	50.0	Less than 0.25	
III	0.25	5.0	0.25	5.0
III-S	0.25	5.0	Less than 0.25	1.0
IV	0.10	1.0	0.10	1.0

Figure 1.2-5. Power limitations by class of station.

Class A, B and D stations may be assigned to these channels
540, 640*, 650*, 660*, 670*, 680, 690, 700*, 710, 720*, 730, 740, 750*, 760*, 770*, 780*, 800, 810, 820*, 830*, 840*, 850, 860, 870*, 880*, 890*, 900, 990, 940, 1000, 1010, 1020*, 1030*, 1040*, 1050, 1060, 1070, 1080, 1090, 1100*, 1110, 1120*, 1130, 1140, 1160*, 1170, 1180*, 1190, 1200*, 1210*, 1220, 1500, 1510, 1520, 1530, 1540, 1550, 1560, 1570, 1580 kHz

* Only one Class A station will be assigned on this channel.

REGIONAL CHANNELS

Class B and D stations may be assigned to these channels
550, 560, 570, 580, 590, 600, 610, 620, 630, 790, 910, 920, 930, 950, 960, 970, 980, 1150, 1250, 1260, 1270, 1280, 1290, 1300, 1310, 1320, 1330, 1350, 1360, 1370, 1380, 1390, 1410, 1420, 1430, 1440, 1460, 1470, 1480, 1590, 1600, 1610, 1620, 1630, 1640, 1650, 1660, 1670, 1680, 1690 and 1700 kHz

LOCAL CHANNELS

Only class C stations will be assigned to these channels
1230, 1240, 1340, 1400, 1450, and 1490 kHz

Figure 1.2-6. Current channel utilization in the United States (Clear, Regional, Local channels).

in the United States had been long established, the station classifications still are primarily related to frequency. The current scheme is shown in Figures 1.2-6 and 1.2-7.

FM BROADCASTING FREQUENCY ALLOCATION

The frequency band 88–108 MHz is allocated for FM broadcasting in Region 2 and, with some exceptions, the same is true in Regions 1 and 3 as well. However, unlike AM, FM broadcast allotments are largely a domestic matter, (especially in large countries such as the United States) due to the limited nature of signal propagation at these frequencies. Although there are some international regulations regarding FM broadcasting, there is no region-wide FM agreement in Region 2. Instead, there are bilateral agreements between the U.S. and Canada and the U.S. and Mexico. Both regulate the use of FM channels in the border areas and specify technical standards in order to insure system compatibility.

The basic plan of the FM broadcast band is shown in Figure 1.2-8. The band is divided into 100 channels, each 200 kHz wide. In the United States, the lower 20 channels, located between 88 and 92 MHz, have been reserved for non-commercial broadcasting, how-

Class of Station	Power in kW Daytime hrs. Minimum	Maximum	Power in kW Nighttime hrs. Minimum	Maximum
A	10.0	50.0	10.00	50.0
B	0.25	50.0	0.25	50.0 (below 1605 kHz)
B	0.25	10.0	0.25	10.0 (1605–1705 kHz)
C	0.25	1.0	0.25	1.0
D	0.25	50.0	–	Less than 0.25

Figure 1.2-7. Power limitations by class of station.

82–88 MHz	88 MHz–92 MHz	92 MHz–108 MHz
CHANNEL 6 TELEVISION	US Non-commercial FM BROADCASTING	US Commercial

Figure 1.2-8. Limits of the FM broadcast band.

ever such stations are not restricted to these channels. In addition, although it is part of television channel 6, the frequency 87.9 MHz can be used for low power non-commercial FM stations, but its use is severely restricted.

As with AM broadcasting, there are different classes of FM stations designed to provide different types of service. As the demand for more stations has increased so has the number of different classes of stations. In response, the FCC has significantly modified the criteria concerning the use of the frequencies. In June 1983, the FCC concluded a lengthy rule making proceeding and modified the domestic allotment criteria for FM broadcasting. Prior to this action, there were three classes of stations in the 80 channel commercial band. Twenty of these channels were used for lower power Class A stations having a maximum effective radiated power (ERP) of 3 kW and a maximum antenna height-above-average-terrain (HAAT) of 300 ft. Class A stations had a 1 mV/m service radius of about 15 miles. Higher power Class B or C stations operated on the remaining 60 channels. Whether a station was designated Class B or C depended on where it was located.[7] Class B stations were located in Zone I or I-A. Zone I is the northeast U.S., extending south to the Virginia-North Carolina border and west to the Mississippi River. Zone I-A is all of California but its northernmost portion, plus Puerto Rico and the U.S. Virgin Islands. Class C stations operated elsewhere in the country, which is referred to as Zone II.

Class B stations operate with a maximum ERP of 50 kW at 500 ft. (152 m) HAAT and have a service radius of about 32 miles. Class C stations operate with a maximum power of 100 kW at 2000 ft. (610 m) HAAT, for a service radius of approximately 57 miles. FM stations in each of these classes may elect to operate at a HAAT above the maximum, but in such cases they are required to make a compensatory reduction in ERP as noted in FCC Rules and Regulations Section 73.211.

This system was changed drastically by the above noted 1983 proceeding. The FCC:

1. Permitted Class A stations to operate on channels previously reserved for Class B or C stations.
2. Created three new classes of FM stations. Class B1 stations are permitted to operate in Zones I and I-A with a maximum ERP of 25 kW at 100 m (328 ft. HAAT); Class C1 and Class C2 stations are permitted to operate in Zone II. Class C1 stations are permitted a maximum ERP of 100 kW at 299 m (981 ft.) HAAT and Class C2 stations are permitted a maximum ERP of 50 kW at 150 m (492 ft.) HAAT.
3. Required stations that were previously licensed as a Class B or C, and were not operating at the minimum level specified for their class under the new rules, to upgrade their facilities within 3 years. Otherwise, the under-minimum facilities would be reclassified to the appropriate lower class based on the facilities they used.
4. Increased the maximum antenna HAAT for Class A stations to 100 m (328 ft.).

[7] The concept of allowing different classes of stations in different areas or "zones" is based on the population density of the areas. The Commission has assumed that there is less need for wide areas of service in areas of dense poplation.

Station Class	Maximum ERP	Maximum HAAT in meters (ft.)	Expected Service Radius
A	6 kW	100 (328)	28 km
B1	25 kW	100 (328)	39 km
B	50 kW	150 (492)	52 km
C3	25 kW	100 (328)	39 km
C2	50 kW	150 (492)	52 km
C1	100 kW	299 (981)	72 km
C	100 kW	600 (1968)	92 km

Puerto Rico and the Virgin Islands

Station Class	Maximum ERP	Maximum HAAT in meters (ft.)	Expected Service Radius
A	6 kW	240 (787)	42 km
B1	25 kW	150 (492)	46 km
B	50 kW	472 (1549)	78 km

Figure 1.2-9. Standards for FM allotments (a) for locations other than Puerto Rico and the Virgin Islands (b).

In 1989, the FCC further modified the rules to permit Class A stations to operate with an ERP of 6 kW and a HAAT of 100 m, (328 ft.) and added an additional classification C3 that permits operation in Zone II with a maximum of 25 kW at 100 m (328 ft.) HAAT.

In addition to the above, there also are Class D stations that operate as non-commercial educational stations with power not in excess of 10 W. However, applications for this class of station are no longer being accepted.[8] A complete list of the station classes and a summary of FM allotment standards can be found in Figure 1.2-9.

Unlike AM broadcasting, where a new station may be applied for at any location where it can meet applicable criteria, the use of commercial FM channels (channels 221–300) is governed by the FM "Table of Allotments" found in Section 73.202 of the FCC's rules. This table lists all FM channel allotments that have been made available for use. Most already are in use. If the Table does not list a vacant channel in the desired community, the prospective applicant must file a rule making petition with the FCC seeking to add such a channel for the community. The rule making petition proposing such addition must provide a show-

[8] Section 73.512(c) FCC Rules and Regulations

Relation	Co-Channel	200 kHz	400/600 kHz	10.6/10.8 MHz
A–A	115(71)	72(45)	31(19)	10(6)
A–B1	143(89)	96(60)	48(30)	12(7)
A–B	178(111)	113(70)	69(43)	15(9)
A–C3	142(88)	89(55)	42(26)	12(7)
A–C2	166(103)	106(66)	55(34)	15(9)
A–C1	200(124)	133(83)	75(47)	22(14)
A–C	226(140)	165(103)	95(59)	29(18)
B1–B1	175(109)	114(71)	50(31)	14(9)
B1–B	211(131)	145(90)	71(44)	17(11)
B1–C3	175(109)	114(71)	50(31)	14(9)
B1–C2	200(124)	134(83)	56(35)	17(11)
B1–C1	233(145)	161(100)	77(48)	24(15)
B1–C	259(161)	193(120)	105(65)	31(19)
B–B	241(150)	169(105)	74(46)	20(12)
B–C3	211(131)	145(90)	71(44)	17(11)
B–C2	241(150)	169(105)	74(46)	20(12)
B–C1	270(168)	195(121)	79(49)	27(17)
B–C	274(170)	217(135)	105(65)	35(22)
C3–C3	153 (95)	99 (62)	43(27)	14(9)
C3–C2	177(110)	117(73)	56(35)	17(11)
C3–C1	211(131)	144(90)	76(47)	24(15)
C3–C	237(147)	176(109)	96(60)	31(19)
C2–C2	190(118)	130(81)	58(36)	20(12)
C2–C1	224(139)	158(98)	79(49)	27(17)
C2–C	249(155)	188(117)	105(65)	35(22)
C1–C1	245(152)	177(110)	82(51)	34(21)
C1–C	270(168)	209(130)	105(65)	41(25)
C–C	290(180)	241(150)	105(65)	48(30)

Figure 1.2-10. Minimum distance separation requirements.

ing that the proposal meets the separation requirements that are applicable to the class of station being proposed. A complete list of the spacing requirements, including those that pertain to stations located near Canada and Mexico, on channel 253 (98.5 MHz) and those located near television stations operating on channel 6, are provided in Figures 1.2-10, 1.2-11, 1.2-12 and 1.2-13. Alternatively, a petitioner can propose to modify the Table by deleting a vacant existing allotment or by changing the frequency of an existing station and thereby achieve compliance with these spacing requirements.

Relation	Co-Channel	200 kHz	400/600 kHz	10.6/10.8 MHz
A–A	132	85	45	8
A–B1	180	113	62	16
A–B	206	132	76	16
A–C1	239	164	98	32
A–C	242	177	108	32
B1–B1	197	131	70	24
B1–B	223	149	84	24
B1–C1	256	181	106	40
B1–C	259	195	116	40
B–B	237	164	94	24
B–C1	271	195	115	40
B–C	274	209	125	40
C1–C1	292	217	134	48
C1–C	302	230	144	48
C–C	306	241	153	48

Note: For purposes of the Canadian Agreement, U.S. Class A allotments with more than 3 kW ERP and an antenna HAAT above 100 meters (328 ft.), or with equivalent lower ERP and higher HAAT, are considered to be Class B1 allotments. Also, U.S. Class C2 allotments are considered to be Class B, and U.S. Class C3 allotments are considered to be Class B1.

Figure 1.2-11. Minimum distance separation requirements in kilometers—Canadian agreement.

Relation	Co-Channel	200 kHz	400/600 kHz	10.6/10.8 MHz
A–A	105(65)	65(40)	25(15)	8(5)
A–B	175(110)	105(65)	65(40)	16(10)
A–C	210(130)	170(105)	105(65)	32(20)
A–D	95(60)	50(30)	25(15)	8(5)
B–B	240(150)	170(105)	65(40)	25(15)
B–C	270(170)	215(135)	105(65)	40(25)
B–D	170(105)	95(60)	65(40)	16(10)
C–C	290(180)	240(150)	105(65)	48(30)
C–D	200(125)	155(95)	105(65)	25(15)
D–D	18(11)	10(6)	5(3)	3(2)

Note: For purposes of the Mexican Agreement, U.S. Class A allotments with more than 3 kW ERP and an antenna HAAT above 91 m (299 ft.), or with equivalent lower ERP and higher HAAT, are considered to be Class B allotments. Also, U.S. Class B1, C2 and C3 allotments are considered to be Class B, and U.S. Class C1 allotments are considered to be Class C.

Figure 1.2-12. Minimum distance separation requirements in kilometers (miles)—Mexican Agreement.

Assignment of stations on the non-commercial educational channels (201–220) is accomplished more in the manner that is followed in AM, where an application includes a showing that interference will not be caused to other stations. In addition, for proposals to use Channels 218, 219 and 220, compliance with applicable separation requirements to any allotments on higher, adjacent commercial channels is required.

In addition to the regular FM broadcast stations, there are two other types of stations that are permitted to operate in the FM band on a secondary basis. These are FM translator stations and FM booster stations. An FM booster station retransmits the signal of a primary station on the primary station's channel in order to serve areas where the primary station's signal is inadequate. An FM translator station is similar to an FM booster station, except that the signal is not retransmitted on the same channel but instead is translated to a different channel. These stations are authorized under Part 74, Subpart L of the FCC Rules, which among other things, requires that such stations provide protection from interference to all full service FM broadcast stations.

TELEVISION BROADCASTING FREQUENCY ALLOCATION

The frequency bands used for television broadcasting in the United States are illustrated in Figure 1.2-14. They include the low-VHF band channels 2 to 4 (54 MHz to 72 MHz) and 5 and 6 (76 MHz to 88 MHz),

FM Class	TV Zone I	TV Zone II and III
A	17	22
B1	19	23
B	22	26
C3	19	23
C2	22	26
C1	29	33
C	36	41

Figure 1.2-13. Minimum distance separation requirements in kilometers to TV channel 6 from FM stations on channel 253 (98.5 MHz) kilometers.

54–72 MHz	76–88 MHz	174–216 MHz	470–806 MHz
TV Channels 2–4	TV Channels 5–6	TV Channels 7–13	TV Channels 14–69

Figure 1.2-14. Limits of the Television broadcast bands.

FM CLASS	TV ZONE I	TV ZONE II and III
A	17	22
B1	19	23
B	22	26
C3	19	23
C2	22	26
C1	29	33
C	36	41

Figure 1.2-16. Minimum distance separation requirements in kilometers to TV channel 6 from FM stations on channel 253 (98.5 MHz).

the high-VHF band channels 7 through 13 (174 MHz to 216 MHz), and the UHF band channels 14 through 69 (470 MHz through 806 MHz). The greater portion of all of these bands is allocated for broadcasting throughout the world; but this allocation is not uniform and in many areas other uses such as land mobile are permitted on a secondary basis. This is also the case in the U.S., where certain UHF television channels are now used for land mobile in some major cities. In general, due to the limited extent of radio wave propagation in the television band, TV allocations, like FM, are basically a domestic matter, with few international regulations. Although no regional agreement exists, the United States does have agreements with both Canada and Mexico concerning television allocation.

Like FM broadcasting, TV allocations in the United States are governed by a "Table of Frequency Allotments." The current FCC Rules (fall 1997) contain two allotment tables one for analog stations in Section 73.606 and one for digital stations in Section 73.622. These tables contain all commercial as well as non-commercial allotments (the latter are identified by an asterisk). Although the FCC is no longer accepting petitions to modify the analog table, a description of how it was constructed is given below for historical purposes. The analog table is based on the co-channel adjacent channel, and FM channel 253 separation criteria contained in Section 73.610 (see Figures 1.2-15 and 1.2-16), and on the UHF channel separation criteria listed in Table II of Section 73.698 of the FCC Rules. As can be seen in Figures 1.2-15 and 1.2-16, there are different requirements for different zones or areas of the country. Zone I is the northeast U.S. extending south to the Virginia-North Carolina border and west to the Mississippi River. Zone II consists of that portion of the United States that is not in Zone I or Zone III, along with Puerto Rico, Alaska, the Hawaiian Islands and the Virgin Islands. Zone III is that portion of the southeast United States extending from the east coast of Georgia westward to the Mexican border. An exact description of the zones is contained in Section 73.609 of the FCC Rules. As with FM, the zones reflect the differing population densities in various parts of the country. In addition, differences in propagation conditions were also considered. The closer spacings in Zone I recognize the fact that in the northeast portion of the United States there are many large population centers needing stations of their own, which are close enough to one another to lessen the need for wide area service. Zone II is characterized by fewer population centers, usually smaller and further apart. For them, wide area service is a necessity. Finally, because the area of the country along the Gulf Coast (Zone III) is susceptible to high levels of tropospheric propagation, stations in that area need to be spaced further apart to minimize interference. Although there are differences in the powers authorized for low-VHF, high-VHF, and UHF stations, unlike AM and FM broadcasting, there are no class designations as such in television. The differing power limitations reflect the differences in signal propagation for low VHF, high VHF and UHF.

The Digital Television (DTV) "Table of Allotments was developed through a complex analysis that took into consideration actual computations of coverage and interference. The DTV Table contains an entry for each analog station that was in operation or for which an application had been filed prior to a cutoff date specified by the FCC.

A petition for rule making may be filed with the FCC seeking to modify the Table to include a new allotment. Such petitions must provide a showing that the proposal complies with applicable separation criteria that serve to prevent mutual interference (see Figure 1.2-17). Furthermore, applications to construct a station on a new allotment must demonstrate that the proposed facility would provide applicable coverage to the city of license and that no new interference would be caused to existing stations either analog or digital.

In addition to the regular television broadcast stations, there are three other types of analog stations that are permitted to operate in the television bands. However, these other stations operate with low power,

ZONE	CHANNELS 2–13	CHANNELS 14–69
I	272.7 (169.5 mi.)	248.6 (154.5 mi.)
II	304.9 (189.5 mi.)	280.8 (174.5 mi.)
III	353.2 (219.5 mi.)	329.0 (204.5 mi.)

MINIMUM ADJACENT CHANNEL SEPARATION DISTANCES IN KILOMETERS (MILES)

Channels 2–13	95.7 km (59.5 mi.)
Channels 14–69	87.7 km (54.5 mi.)

In addition to the above there are other constraints on UHF channels due to various types of problems caused by the mixing of the signals of stations operating on different channels. A complete table of these restrictions, referred to as the "UHF Taboos", is contained in Section 73.698 of the FCC Rules.

Figure 1.2-15. Minimum co-channel television separation distances in kilometers (miles).

VHF Channels 2–13	DTV to DTV	DTV to Analog TV
Co-channel—Zone I	244.6 (152.0)	244.6 (152.0)
Co-channel—Zone II & III	273.6 (170.0)	273.6 (170.0)
Adjacent Channel—Zone I	<20.0 (12.4) or >110.0 (68.4)	<9.0 (5.6) or >125.0 (77.7)
Adjacent Channel—Zone II & III	<23.0 (14.3) or >110.0 (68.4)	<11.0 (6.8) or >125.0 (77.7)
UHF Channels	**DTV to DTV**	**DTV to Analog TV**
Co-channel—Zone I	196.3 (122.0)	217.3 (135.2)
Co-channel—Zone II & III	223.7 (139.0)	244.6 (152.0)
Adjacent Channel—Zone I	<24.0 (14.9) or >110.0 (68.4)	<12.0 (7.5) or >106.0 (65.9)
Adjacent Channel—Zone II & III	<24.0 (14.9) or >110.0 (68.4)	<12.0 (7.5) or >106.0 (65.9)
[9]Taboo Channels—Zone I		<24.1 (15.0) or > 80.5 (50.0)
[10]Taboo Channels—Zone II & III		<24.1 (15.0) or > 96.6 (60.0)

[9] Values are only specified for DTV to analog TV and are for DTV channels ±2, ±3, ±4, ±7, ±8, +14 and +15 channels away from the analog TV channel.
[10] Ibid.

Figure 1.2-17. Required television separation distances in kilometers (miles)—DTV to analog TV and DTV to DTV.

on a secondary basis. These are Low Power Television Stations (LPTV), TV Translator Stations and TV Booster Stations. A TV Booster Station retransmits the signal of a primary station on the primary station's channel in order to serve areas where the primary station's signal is inadequate. A TV Translator Station is similar to a TV Booster Station, except that the signal is not retransmitted on the same channel but instead is translated to a different channel. Low Power Television Stations may retransmit the signals of another station or they may originate programming. All of these stations are authorized in accordance with Part 74, Subpart G of the FCC Rules, which requires full protection from interference to all regular television broadcast stations. In addition, these stations are not accorded any protection from the operation of regular television broadcast stations either analog or digital. As of this writing, the future of LPTV, TV Translator and TV Booster Stations is unclear since the advent of DTV will displace many of these stations since they were not protected in the DTV planning process.

AUXILIARY BROADCAST SERVICES FREQUENCY ALLOCATION

Although all auxiliary broadcast services share a common role in support of AM, FM or TV broadcast operations, there are important differences between them. The nature of the service they provide varies as does the frequency band in which they operate. For these reasons, the allotments available for each auxiliary service need to be discussed separately. However, before turning to a specific discussion of each service, some general comments are necessary.

The steady growth in the number of AM, FM and TV stations including DTV, as well as their desire to use more advanced technology in every aspect of their operation, has greatly increased demand for spectrum in all auxiliary services. Because of the continuing demand for spectrum from a multitude of other broadcast and non-broadcast activities, it has been difficult for the FCC to allocate more spectrum to alleviate the congestion faced by the auxiliary services. At the same time, some relief has come through changes in the FCC rules to facilitate the use of newer, more spectrum-efficient, technology. In using this material, the reader should be aware that additional changes may take place. Accordingly, the careful reader should check the current regulations governing each particular radio service.

Recognizing that available spectrum is limited, it is imperative for broadcasters to use it efficiently. To do this, it is first necessary to understand what spectrum is allocated, how it may be used and what advantages or disadvantages, if any, may be involved in the use of a particular band for each of the auxiliary services discussed below.

Remote Pickup Broadcast Stations (RPUs)

RPUs are mobile or portable facilities used to transmit live on-the-air programming from a temporary remote location, such as a shopping center or football game, to the station's studio facilities. This material can be taped for later rebroadcast, or it can be incorporated into actual on-going live broadcasts. Radio stations typically have several RPUs that may be licensed to one or more frequencies.

In November 1984, the FCC significantly revised its radio broadcast auxiliary frequency allocations to permit operational use of narrow band technologies in the broadcast remote pickup service. The Commission's goal was to foster spectrum efficiency in a flexible manner. Broadcasters and equipment manufacturers who wanted to operate narrow band equipment, the Commission believed, should not be precluded from doing so by rigid FCC rules. However, these changes have not yet been formally implemented by the FCC due to delays in setting up the mechanism to handle the expected applications, but formal implementation is expected in the near future. In the meantime, the FCC will accept applications based on this revised allocation.

The following are the frequency allocations for radio broadcast remote pickup stations.

1. 25.85–26.48 MHz: There are a total of 25 frequencies available in this band for use by remote pickup broadcast stations. Bandwidth is limited to 20 kHz except between 25.87 and 26.03 MHz where 40 kHz is permitted. Please note that the use of the frequencies between 25.87 and 26.09 are subject to the condition that no harmful interference is caused to broadcast stations sharing this band. It is also noted that the frequencies between 26.100 and 26.175 have been allocated on a worldwide basis

to the Maritime Mobile Services. Although this service has not yet been implemented, at such time as the maritime service is implemented, these frequencies may no longer be available for use by remote pickup stations. It is further noted that the frequencies between 26.175 and 26.47 are allocated on a worldwide basis for use by various types of fixed and mobile operations. Hence, in selecting a frequency in this band for remote pickup use, the implications of the above should be taken into consideration.

2. 152.855–153.3645 MHz: There are a total of 54 frequencies available in this band for use by remote pickup broadcast stations. Each channel is 5 kHz wide, and they may be stacked to form a single channel with a maximum bandwidth of 30 kHz. This band is shared with the Private Land Mobile Radio Service and the Maritime Service, and operation of remote pickup stations is subject to the condition that no harmful interference is caused to these other services. Please note that these frequencies are not available to network entities or for use on board aircraft unless they were previously licensed for such purpose.
3. 160.86–161.4 MHz: There are a total of 108 frequencies available in this band for use by remote pickup broadcast—but they are available *only* to stations in Puerto Rico or the Virgin Islands where they are shared with the Public Safety and Land Transportation Radio Services. Each channel is 5 kHz wide and they may be stacked to form a single channel width a maximum bandwidth of 30 kHz.
4. 161.625–161.775 MHz: There are a total of 30 frequencies available in this band for use by remote pickup broadcast stations. Each channel is 5 kHz wide and they may be stacked to form a single channel width a maximum bandwidth of 30 kHz. These frequencies are *not* available to network entities and are *not* available for use in Puerto Rico or the Virgin Islands. Also, Public Safety and Land Transportation Radio Service stations may continue to operate on these frequencies on a non-interference basis.
5. 166.25 and 170.15 MHz: These frequencies may be used by remote pickup stations with a maximum bandwidth of 25 kHz; however, the area in which they may be used is restricted. A description of the area in which they may be used is found in FCC Rule Section 74.402.
6. 450.005–450.025 MHz, 450.975–450.995 MHz, 455.005–455.025 MHz and 455.975–455.995 MHz: These frequencies may be used by remote pickup stations only for the transmission of operational communications, including tones for signaling and for remote control and automatic transmission system control and telemetry. Bandwidth is limited to a maximum of 10 kHz.
7. 450.025–450.625 MHz and 455.025–455.625 MHz: There are a total of 240 frequencies available in these bands for use by remote pickup broadcast stations. Each channel is 5 kHz wide, and they may be stacked to form a single channel with a maximum bandwidth of 50 kHz.
8. 450.625–450.875 MHz and 455.625–455.875 MHz: There are a total of 20 frequencies available in these bands for use by remote pickup broadcast stations. Each channel is 25 kHz wide, and they may be stacked to form a single channel with a maximum bandwidth of 50 kHz. Users who commit to 50 kHz bandwidths and are transmitting program material have primary use of these channels.
9. 450.875–450.975 MHz and 455.875–455.975 MHz: There are four frequencies available in these bands for use by remote pickup broadcast stations. Each channel is 50 kHz wide, and they may be stacked to form a single channel with a maximum bandwidth of 100 kHz. Users committed to 100 kHz bandwidths to transmit program material will have primary use of these channels.

Aural Broadcast Auxiliary Stations

Aural Broadcast Auxiliary Stations include studio transmitter link (STL), intercity relay (ICR) and Microwave Booster Stations used by radio broadcast stations. STL stations are fixed stations used for transmitting program material between the studio and the transmitter of a broadcasting station. ICR stations are fixed stations used for the transmission of program material between broadcasting stations, except international stations, for simultaneous or delayed broadcast. ICRs may also be used on a secondary basis to transmit program material from FM stations to co-owned FM translator or booster stations. Microwave Booster Stations are used to relay the signals of a STL or ICR station over a path that cannot be covered with a single station. They receive and transmit on the same frequency. One or more Microwave Booster Stations may be authorized to licensees of STLs or ICRs. It should be noted that stations in the Aural Broadcast Auxiliary service may be authorized on a secondary non-interference basis to licensees of TV broadcast stations to transmit aural material.

The following frequencies are available for assignment to STL, ICR and Microwave Booster Stations:

1. 942–944 MHz: These frequencies are available for assignment *only* in Puerto Rico. Stations licensed on these frequencies in other parts of the United States prior to November 21, 1984, may continue to operate on a co-equal primary basis in this band.
2. 944–952 MHz: There are 320 channels (25 kHz wide) available in this band. The channels may be stacked to form a single channel up to 300 kHz wide. Separately, stations also may be authorized additional 25 kHz wide channels up to a grand total of 20 channels. The use of these frequencies by ICR stations is subject to the condition that no harmful interference is caused to other classes of stations.
3. 18760–18820 MHz and 19100–19160 MHz: There are 24 channels (5 MHz wide) available in these bands. These frequencies are shared on a co-pri-

mary basis with other fixed services and their use is subject to the rigorous coordination requirements of FCC Rule Section 21.100(d).

4. 31000–31300 MHz: Frequencies in this band are shared on a co-equal basis with stations in the fixed and mobile services operating under parts 21, 78, and 95 of the FCC rules. No interference protection is afforded to fixed or mobile stations operating in this band.

Television Broadcast Auxiliary Stations

The demand for the spectrum allocated for television auxiliary services is significantly greater than for spectrum allocated for radio broadcasting services. In addition to the extensive local demand, network remote units travel extensively and compete with local broadcasters for available frequencies, resulting in increased spectrum congestion. Finally, it is also important to recognize that the variety of activities undertaken by television broadcasters usually requires more complex auxiliary systems than is the case in the radio industry.

The following are the types of Television Broadcast Auxiliary Stations: TV Pickup Stations, TV STL stations (studio transmitter link), TV Relay Stations, TV Translator Relay Stations and TV Microwave Booster Stations.

TV Pickup Stations are land mobile stations used for the transmission of TV program material and related communications from the scenes of events to TV broadcast or Low Power TV (LPTV) stations.

TV STL Stations are fixed stations used for the transmission of program material and related communications from the studio to the transmitter of a TV broadcast or LPTV station.

TV Relay Stations are fixed stations used for transmitting visual program material between TV broadcast or LPTV stations or for the relay of transmissions from a remote pickup station to a single TV station.

TV Translator Relay Stations are fixed stations used for relaying programs and signals of TV broadcast stations to LPTV stations or TV translator stations.

TV Microwave Booster Stations are fixed stations used to receive and amplify signals of TV Pickup, TV STL, TV Relay or TV Translator Relay Stations and retransmit them on the same frequency. These stations are used to transmit signals over a path that cannot be covered by a single transmitter.

The following bands are available for assignment to TV Pickup Stations:

1. 1990–2110 MHz: Seven channels are available in this band which is also available for assignment to all the other types of Television Broadcast Auxiliary Stations. (*Note:* As this *Handbook* goes to press the FCC is considering reallocating some of this band in one of its open rule making proceedings.)
2. 2450–2483.5 MHz: Two channels are available in this band which is also available for assignment to all the other types of Television Broadcast Auxiliary Stations. This band is shared with industrial, scientific and medical (ISM) devices and is not afforded any protection from these ISM devices.
3. 6425–6525 MHz: The channels available in this band are for mobile use only and are co-equally shared with mobile stations licensed under Parts 21, 78 and 94 of the FCC Rules. The available channel bandwidth varies from 1 MHz to 25 MHz and Section 74.602(j) of the FCC Rules contains further explanation concerning the usage of this band.
4. 6875–7125 MHz: Ten channels 25 MHz wide, are available in this band which is also available for assignment to all the other types of Television Broadcast Auxiliary Stations.
5. 12700–13250 MHz: This band contains 43 channels; however, the channels overlap and therefore if use of this band is contemplated Section 74.602 of the FCC Rules should be consulted for a more complete understanding of their usage. This band is also available for assignment to all other types of Television Broadcast Auxiliary Station.
6. 38.6–40 GHz: This band is available without channel bandwidth limitation on a secondary basis to fixed stations.

The following are available for assignment to TV STL, TV Relay or TV Translator Relay Stations.

1. 1990–2110 MHz: Seven channels are available in this band which is also available for assignment to TV Pickup Stations.
2. 2450–2483.5 MHz: Two channels are available in this band which is also available for assignment to TV Pickup Stations.
3. 6875–7125 MHz: Ten channels 25 MHz wide, are available in this band which is also available for assignment to TV Pickup Stations.
4. 12700–13250 MHz: This band contains 43 channels; however, the channels overlap and therefore if use of this band is contemplated Section 74.602 of the FCC Rules should be consulted for a more complete understanding of their usage. This band is also available for assignment to TV Pickup Stations. In addition, the channels between 13150 and 13200 MHz are not available within 50 km of the top 100 markets.
5. 17700–18580 MHz and 19260–19700 MHz: Frequencies in these bands are shared on a co-equal basis with stations in the fixed service authorized by Parts 21, 78 and 94 of the FCC Rules. The available channel bandwidth varies from 2 MHz to 80 MHz. Complete details concerning the use of this band are contained in Section 74.602 of the FCC Rules.

In addition to the above frequencies, TV STL and TV Relay Stations also may be authorized to use UHF-TV channels 14–69 on a secondary basis provided no interference is caused to TV and LPTV stations operating in this band.

Furthermore, the aural portion of television broadcast program material may be transmitted over an aural broadcast STL or ICR station on a secondary, non-interference basis. Likewise, remote pickup stations may be used to transmit the aural portion of television program material.

CONCLUSION

As can be seen from the material contained in this chapter, frequency allocation is a very complex matter and is subject to frequent changes. Therefore, the reader is advised to consult the FCC Rules and Regulations for a complete description of the current allocation situation, including the procedures and policies currently being applied by the FCC concerning a particular band.

APPENDIX A

ITU Study Groups

Study Group	Subject
1.	Spectrum management
3.	Radiowave propagation
4.	Fixed-satellite service
7.	Science services
8.	Mobile, radio determination, amateur & related satellite services
9.	Fixed service
10.	Broadcasting service (sound)
11.	Broadcasting service (television)

Currently there are no study groups 2, 5 or 6. The numbering scheme reflects that which was in existence before the ITU reorganization.

1.3 FCC ORGANIZATION AND ADMINISTRATIVE PRACTICES

BY THE OFFICE OF PUBLIC AFFAIRS, FEDERAL COMMUNICATIONS COMMISSION, WASHINGTON, DC

STATUTORY AUTHORITY

Congress, through adoption of the Communications Act of 1934 created the Federal Communications Commission (FCC) as an independent regulatory agency. Section I of the Act specifies that the FCC was created: "For the purpose of regulation of interstate and foreign commerce in communication by wire and radio so as to make available, so far as possible, to all the people of the United States a rapid, efficient, nationwide, and worldwide wire and radio communication service with adequate facilities at reasonable charges, for the purpose of the national defense, for the purpose of promoting the safety of life and property through the use of wire and radio communication, and for the purpose of securing a more effective execution of this policy by centralizing authority heretofore granted by law to several agencies and by granting additional authority with respect to interstate and foreign commerce in wire and radio communication. . .".

The Commission

The FCC is directed by five Commissioners appointed by the President, by and with the advice and consent of the Senate for staggered five-year terms. No more than three can be members of the same political party. The President designates one Commissioner as Chairman. The Commissioners make their decisions collectively by formal vote although authority to act on routine matters is normally delegated to the staff.

FCC Organization

The staff of the FCC performs the day-to-day functions of the agency, including: 1) license and application processing, 2) drafting of rulemaking items, 3) enforcing rules and regulations, and 4) formulating policy. Normally, broadcasters deal with the Mass Media Bureau (MMB); however, actions by other elements of the agency may directly affect broadcasters. The Commission reorganized itself in 1995 to establish two new bureaus—Wireless Telecommunications and International—to reflect the changes in the industries it regulates. The organizational units of the FCC are shown in Figure 1.3-1. The staff is divided along functional lines into six operating bureaus and ten offices, which provide support services.

Office of the Managing Director (OMD)

The Managing Director serves as the Chief Operations and Executive official, as supervised and directed by the Chairman. The OMD recommends, to the Chairman, program priorities, resources and position allocations, management and administrative policies. The OMD operates the agency's personnel office and has responsibility for emergency communication policies.

Office of Public Affairs (OPA)

The OPA is responsible for informing the press and public of the FCC's actions, facilitating public participation in the FCC's decision-making processes, and operating many of the FCC's public reference rooms and the library. The OPA issues daily news releases, Public Notices and other informational material; prepares the Annual Report and other publications; and handles telephone, written and walk-in requests for information. The OPA maintains the FCC's Internet homepage at http://www.fcc.gov.

The FCC is expanding its Internet communications pipeline to meet increased use by the public. The OPA also operates a phone fax-on-demand system (202) 418-2830 to allow quick access to forms, press releases, public notices, and other information.

Office of Inspector General (OIG)

The OIG was created by the Inspector General Amendments Act of 1988. The Inspector General conducts and supervises audits and investigations relating to the programs and operations of the agency. The Inspector General recommends policies for activities designed to promote economy, efficiency, and effectiveness, as well as to prevent and detect fraud and abuse in agency programs. The OIG also provides a means for keeping the Chairman, Commissioners and the Congress fully informed about problems and deficiencies at the agency.

Office of Administrative Law Judges (OALG)

Administrative Law Judges preside over hearings and issue initial decisions. Review of initial decisions is done by the full Commission.

Office of General Counsel (OGC)

The General Counsel serves as the chief legal advisor to the Commission and the various bureaus and offices. The General Counsel also represents the Commission before the federal courts of appeals, recommends decisions in adjudicatory matters before the Commission, assists the Commission in its decision making capacity, performs a variety of legal functions

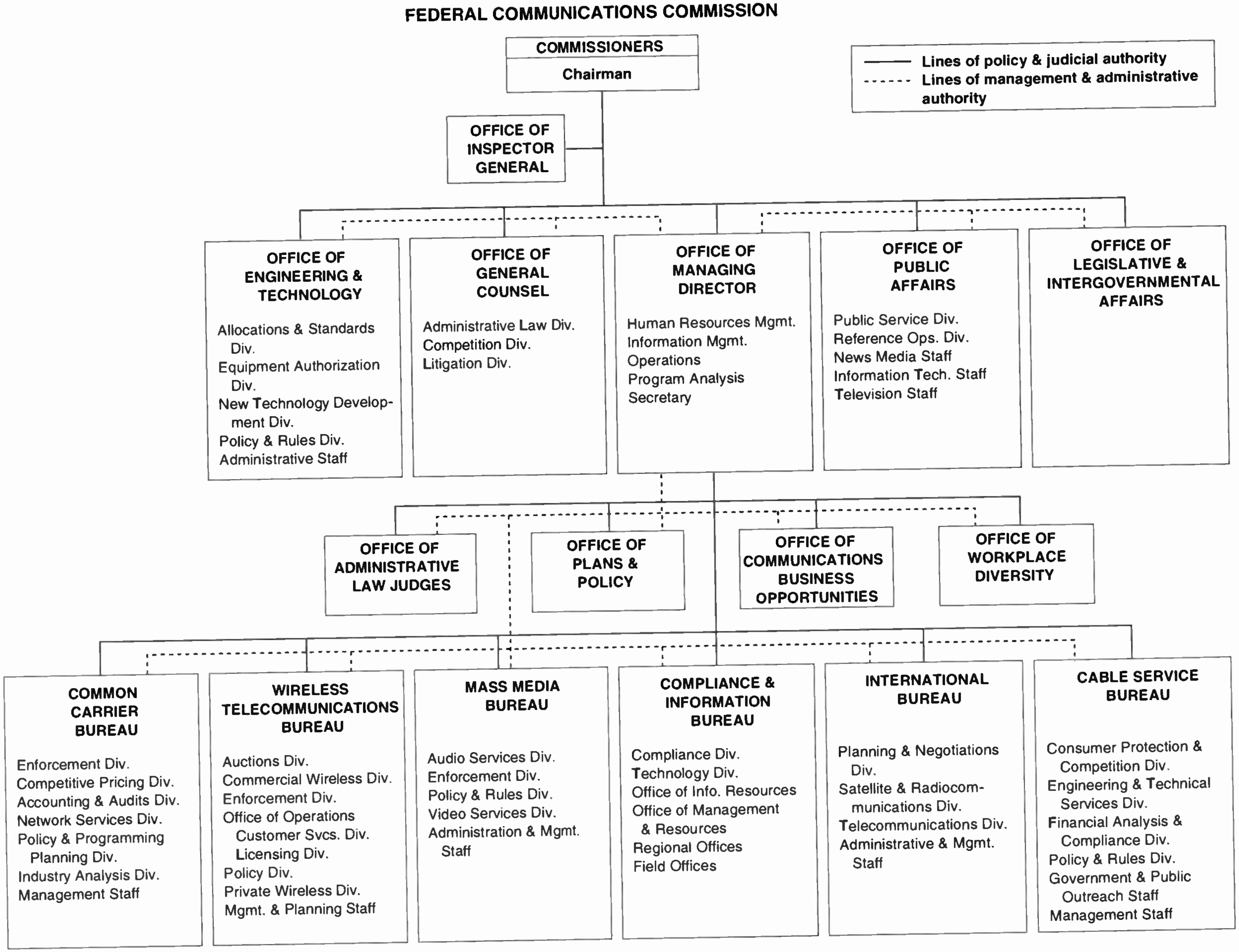


Figure 1.3-1. Operational/organizational chart for the Federal Communications Commission.

regarding internal administrative matters and advises the Commission on fostering competition and promoting deregulation in a competitive environment.

Office of Workplace Diversity (OWD)

The OWD, established in 1995, serves as the principal advisor to the Chairman and Commission on all aspects of workforce diversity, affirmative recruitment, equal employment opportunity, and civil rights within the Commission. The office develops, coordinates, evaluates, and recommends to the Commission internal policies, practices, and programs designed to foster a diverse workforce and to promote equal opportunity for all employees and applicants for employment.

Office of Communication Business Opportunities (OCBO)

The OCBO, established in 1995, is responsible for providing advice to the Commission on issues and policies concerning opportunities for ownership and contracting by small, minority and women-owned communications businesses. The office also advises the Commission on policies to foster equal employment opportunity in the communications industries for minorities, women, and people with disabilities. The office works with entrepreneurs, industry and public interest organizations and individuals to provide information about policies to promote ownership and employment opportunities in the communications industry.

Office of Plans and Policy (OPP)

The OPP serves as the principal economic, technical, and policy adviser to the Commission, analyzing agenda items and developing long term policy on issues that cut across traditional industry and institutional boundaries. The OPP also produces working papers on major policy issues. The continuing focus of OPP's work is to encourage economic growth; enhance access to telecommunications service offerings for consumers, producers, and new entrants, encourage the creation of new services; and promote market-oriented approaches to spectrum management.

Office of Legislative and Intergovernmental Affairs (OLIA)

The OLIA serves as the Commission's principle point of contact with Congress and other governmental entities. Among its functions, the OLIA informs the Congress of the Commission's regulatory decisions, facilitates responses to congressional inquiries, drafts Commission legislative proposals and bill comments, and helps the Commission prepare for congressional hearings. In addition, the OLIA acts on FCC policies and acts as liaison to other federal, state and local government agencies.

Office of Engineering and Technology (OET)

The OET is responsible for managing the non-government use of the spectrum. The OET makes recommendations to the Commission on how the radio spectrum should be allocated and establishes the technical standards to be followed by users. The OET provides technical leadership to create new opportunities for competitive technologies and services for the American public. The OET has four divisions.

Allocations and Standards Division: conducts engineering, economic, legal, social, and statistical studies related to spectrum management. It supports the implementation of new services and technologies by developing technical rules and standards.

Equipment Authorizations Division: authorizes radio transmitters and certain electronic equipment to ensure compliance with the FCC technical standards and to prevent radio interference.

New Technology Development Division: issues licenses for radio experiments, conducts laboratory analyses, and special tests to evaluate performance and spectrum efficiency of new services and technologies.

Policy and Rules Division: prepares rulemaking items on spectrum allocations, radio frequency devices, and industrial, scientific, and medical equipment. It also advises on technical, policy, and standards issues for the various radio services.

BUREAUS

The FCC is organized into six operating bureaus—Common Carrier, Wireless Telecommunications, Compliance and Information, Cable Services, International and Mass Media.

Common Carrier Bureau (CCB)

The CCB is responsible for FCC policies concerning telephone companies that provide interstate telecommunications services to the public through the use of wire-based transmission facilities. These companies provide voice, data, and other transmission services. Under the Communications Act of 1934, as amended by the Telecommunications Act of 1996, the FCC is charged with ensuring that common carriers provide services at just, reasonable, and affordable prices and in a nondiscriminatory manner. The CCB implements FCC regulations and programs designed to meet these objectives. The CCB also recommends and implements proposals designed to encourage the introduction and advancement of competition in telecommunications markets.

Wireless Telecommunications Bureau (WTB)

The WTB oversees the use of radio spectrum to fulfill the communications needs of businesses, local and state governments, public safety service providers, aircraft and ship operators, and individuals. In addition to licensing commercial providers of wireless services, the WTB monitors the more than 2,500,000 licensees that use private wireless radio for personal convenience, to promote safety of life and property, to increase commercial productivity, and to advance the science of telecommunications. The principal functions of the WTB include evaluating new technologies (with the OET); assessing utilization levels of the spectrum and the competitiveness of markets; identifying and allocating spectrum appropriate for licensing; developing through rulemaking the operating rules for radio-based services; assigning licenses through competitive bidding (auctions) and other methods; developing methodologies for conducting complex auctions of radio spectrum; maintaining an antenna structure registration program; licensing commercial radio operators; and enforcing the Commission's rules.

Compliance and Information Bureau (CIB)

The CIB is one of the FCC's primary points of contact with the public. Through its headquarters, staff and various field offices, the CIB informs the public about FCC regulations, policies, practices, and procedures. The CIB also informs the Commissioners and other Bureaus about telecommunications problems and the needs of users; ensures compliance with FCC rules; and uses its technical expertise to solve problems in the communications environment. The Bureau carries out its programs to serve the public and the Commission through 16 district offices, 9 resident agent offices, and 14 remotely-controlled monitoring and radio direction-finding sites located throughout the United States.

Field office activities are directed by three regional offices. The Bureau's radio direction-finding network is controlled from its Columbia, MD, Operations Center. The CIB also maintains a Communications and Crisis Management Center at its Washington, DC, headquarters. The Bureau's field offices are the vital link with the public, licensees, and other users of the radio spectrum. Program and policy development for the Bureau is the responsibility of the Bureau's Divisions and other Offices located at the FCC's headquarters in Washington, DC. In furtherance of the regulatory reform goals of the Telecommunications Act of 1996, the CIB undertook rulemaking proceedings to reduce economic burdens on ship owners and the Commission by privatizing ship inspections previously performed by the Commission on large cargo and small passenger ships and Great Lakes vessels.

Cable Services Bureau (CSB)

The CSB administers and enforces cable television rules and licenses private microwave radio facilities used by cable systems. The Bureau is charged with implementing the 1992 Cable Television and Consumer Protection and Competition Act, which directed the Commission to ensure good service and reasonable rates for cable television subscribers nationwide. The 1996 Telecommunications Act revised many provisions of the 1992 Cable Act and significantly altered the workload for the Bureau. The 1996 Act continues the Bureau's—and the Commission's—process of moving towards competition and away from regulation. The 1996 Act changed the rules of this lane of the information highway for the industry and for consumers. For example, the 1996 Act changed the way cable television subscribers can complain about rates. Prior to the signing of the 1996 Act by President Clinton on February 8, 1996, cable subscribers who had a complaint about rates on the cable programming service tier could file a complaint directly with the Commission. Now, under the 1996 Telecommunications Act, a subscriber must first file with the local franchise authority in the area in which the cable system is located. The Bureau provides a single point of contact for consumers, community officials and the industry for cable related issues pending before the Commission.

International Bureau (IB)

The mission of the IB to promote innovative, efficient, reasonably priced, widely available, reliable, timely, and high-quality domestic and global communications services. The Bureau develops, recommends and administers policies, standards, procedures, and programs for the authorization and regulation of international telecommunications facilities and services and the licensing of domestic and international satellite systems. The Bureau advises and recommends to the Commission, or acts for the Commission under delegated authority, in the development and administration of international telecommunications policies and programs.

Mass Media Bureau (MMB)

The MMB advises the Commission on policy pertaining to broadcasting—television and radio—as well as Multipoint Distribution Service (MDS) (sometimes called *wireless cable*) and Instructional Television Fixed Service (ITFS), a service used mainly by educational entities to provide classroom instruction to multiple locations. The Bureau issues licenses (authorization of service), performs policy and rulemaking functions, and administers the enforcement program for all mass media services. The Bureau has four divisions:

Audio Services Division (ASD): reviews and takes action on all applications filed by the nation's 12,000 FM and AM stations and 2,500 FM translator and booster stations. In March 1996, ASD launched its homepage on the Internet, providing broadcasters with direct access to application status reports, engineering databases, and staff letter decisions of general interest. The Division also has a customer service team and renewal information line, which provide broadcasters and the public with a ready source of information on the status of applications, filing procedures and radio rules.

Enforcement Division: responds to complaints, conducts investigations concerning broadcast stations, handles matters concerning political broadcasting, enforces the broadcast and cable television equal employment opportunity laws and rules, and participates in formal adjudicative proceedings involving broadcast stations. During FY96 the Enforcement Division fully implemented improvements effected in FY95 to streamline its organization and improve speed of service. As a direct result, it was able to expedite enforcement clearances for grant of the increased number of stations changing hands because of the 1996 Telecom Act's liberalization of the multiple ownership rules.

Policy and Rules Division (PRD): conducts rulemaking proceedings which affect the mass media services and provides legal and technical analyses of these rules. The PRD also conducts economic studies to formulate and evaluate Commission policies in terms of their economic effects on the mass media or on other industries or society, as appropriate, and evaluates economic and other studies submitted to the Commission in connection with various proceedings.

Video Services Division (VSD): reviews and takes action on all applications filed by the nation's 1,550 commercial and non-commercial educational television and the nearly 6,850 TV translator and low power television broadcast stations. The VSD is also responsible for the disposition of applications and other requests regarding the development and licensing of facilities in the Multipoint and/or Multichannel Distribution Service (MDS) and the Instructional Television Fixed Service (ITFS).

THE LICENSING PROCESS

Any qualified citizen, company, or group may apply to the FCC for authority to construct a standard (AM), frequency modulation (FM), or television (TV) broadcast station. Licensing of these facilities is prescribed by the Communications Act of 1934, as amended, which sets up certain basic requirements. In general, applicants must satisfy the Commission that they are legally, technically, and financially qualified, and that operation of the proposed station would be in the public interest. Full details of the licensing procedure and station operation are in Part 1 of the Commission's Rules, "Practice and Procedure," and Part 73, "Radio Broadcast Services."

Part 73 includes technical standards for AM, FM, and TV stations, and TV and FM channels (frequency) assignments by states and communities. Copies of the complete rules may be purchased from the Superintendent of Documents, Government Printing Office, Washington, D.C. 20402; (202) 512-1800.

Most applicants retain engineering and legal services in preparing their applications. The Commission does not perform technical or other special studies for prospective applicants nor does it recommend individual lawyers or engineers. Names of firms and individuals practicing before the Commission are listed in various trade publications. (*Broadcasting & Cable Yearbook* and the *Television and Cable Factbook* are two such publications.) These publications can be found in most public libraries or may be obtained by writing: Reed Reference Publishing Company, Attention: Customer Service, 121 Chanlon Road, New Providence, NJ 07974; (1-800-521-8110 and Warren Publishing, Inc., 2115 Ward Court, NW, Washington, D.C. 20037; (202) 872-9200).

The following is a summary of the consecutive steps to be followed in applying for the authorization to build and operate a broadcast station. The application procedure is substantially the same whether the facility sought is AM, FM, or TV.

AM Stations

An applicant must make their own search for an AM frequency on which to operate without causing or receiving interference from existing stations and stations proposed in pending applications. AM stations operate on frequencies from 540 kHz to 1700 kHz—though the frequencies from 1610–1700 are not at present available to new applicants. In the United States and in Region 2, the channels are spaced at 10 kHz intervals. Stations are designated channel by channel to serve various size areas and operate on *clear, regional,* and *local* channels. Dominant clear channel stations (Class A) operate at a maximum power of 50,000 watts; secondary clear channel stations (Class B) operate at a power between 250 watts and 50,000 watts. Many stations operate as daytime-only stations (Class D). (Applications are no longer being accepted for daytime-only stations). Class B stations operating on regional channels may now operate at power levels between 250 watts and 50,000 watts. Class C stations operate on local channels serving limited areas and operating with no more than 1,000 watts day and night. Submission of an application for an AM station requires the payment of an application filing fee.

FM Stations

The FM service includes the frequencies from 88.1 MHz through 107.9 MHz. These frequencies are also known as *FM broadcast channels* (See 47 CFR Section 73.201). The FM channels run from Channel 201 (88.1 MHz) through Channel 300 (107.9 MHz). Channels 221 through 300 (92.1 MHz through 107.9 MHz) are generally used for commercial FM stations. Any applicant wishing to apply for a commercial FM station (or a noncommercial educational station on Channels 221 to 300) must first determine if there are any unused frequencies ("allotments") assigned to the community to be served. The FCC maintains a Table of Allotments (47 CFR Section 73.202) that is helpful in identifying any available channels.

These tables are available from the FCC's Copy Contractor, ITS at 1231 20th Street, NW, Washington, D.C. 20036; (202) 857-3800. (Information on Table of Allotments is also available on the Commission's Internet site at http://www.fcc.gov/mmb/prd/allocate. html.) If an applicant identifies an unused channel that is listed as a first come, first served allotment, then the applicant may file an application (FCC Form 301) for the vacant allotment. If the applicant identifies an unused channel that has a window filing period for the allotment, the applicant may file an application for that channel before the end of the window filing period. If the current FM Table of Allotments indicates that there are no unused channels allotted to the community, the applicant may have a private consulting engineer, of the applicant's choice, perform an engineering study to identify a channel that may be allotted to the desired community.

If the applicant identifies a channel that could be allotted to serve the community, the applicant may submit a petition for rulemaking to the FCC's Allocations Branch (Policy and Rules Division, Mass Media Bureau) requesting allotment of the channel to serve the desired community. If the FCC approves the petition, the channel will be added to the FM Table of Allotments and applications may be filed during a "window" filing period (usually a 30 day period) specified by the FCC. (See *FM Commercial Application Filing Windows* for additional information.)

Applicants filing for a commercial FM station during a window filing period, or filing for a first come, first served vacant allotment, must include the application filing fee with the construction permit application. Channels 201–220 (88.1 MHz to 91.9 MHz) are reserved for noncommercial educational operation. Applicants for noncommercial FM radio stations must first conduct an engineering study for a suitable channel within the reserved band to serve the desired community.

Once the engineering study has been completed and a suitable channel within the reserved band is identified to serve the desired community, an application (FCC Form 340) may be filed for that channel. The application is then placed on an *A* cutoff list, which defines a date by which any competing applications or petition to deny must be filed. If no competing applications are filed, a legal review and further technical review is performed on the application, with that application being granted if it is acceptable. Channels in the unreserved portion of the FM band (Channel 221 through Channel 300 or 92.1 MHz–107.9 MHz) may be used for noncommercial programming. However, procedures governing commercial FM channels will apply (i.e., window filing periods, technical requirements). No application filing fee is required for applicants seeking authorization for a noncommercial educational station. The Commission authorizes seven classes of FM stations: A, B1, B, C3, C2, C1, and C.

The maximum power for a Class A station is 6,000 watts, Class B1 is 25,000 watts, Class B 50,000 watts, Class C3 is 25,000 watts, Class C2 is 50,000 watts, Class C1 is 100,000 watts, and Class C is 100,000 watts but at a higher (300 to 600 meter) antenna height above average terrain (AHAAT). The minimum power authorized for any FM station is 100 watts. (*See Low Power Broadcast Radio Stations.*)

FM Translator Stations

FM translators comprise a low-power service on the FM broadcast band (88 to 108 MHz) which complements the primary FM service. This service was created in 1970 to allow FM stations to provide supplementary service to areas in which direct reception of radio service is unsatisfactory due to distance or intervening terrain barriers. Translators may not originate programming, except for very limited fund raising efforts. Translator stations rebroadcasting commercial FM stations may be authorized on Channel 221 through 300 (92.1 MHz to 107.9 MHz), while translators rebroadcasting a noncommercial educational FM station may be authorized on any FM channel (201 to 300). The maximum effective radiated power permitted for any translator station is 250 watts. The maximum power permitted for a particular transmitter station may be further limited by the location of that station. Applications for a construction permit for an FM translator station must be filed on FCC Form 349. Once construction is completed in accordance with the construction permit granted by the FCC, a license application on FCC Form 350 is then necessary. An application filing fee is required for FM translator stations which rebroadcast a commercial FM primary station. Further information about FM translator stations is available in the Fact Sheet, *FM Translator and Booster Stations,* available from the Commission's Public Service Division.

TV Stations

As is the case in applying for a commercial FM station, an applicant for a TV station must request an unused channel listed in the Commission's Table of Allotments. Television channels are designated commercial or noncommercial. If there are no channels assigned to the city where an applicant proposes to construct the new station, the applicant may have a private engineer do a channel search. Once a channel is identified, a petition for rulemaking to add the channel to the Table of Allotments may be filed with the Allocations Branch. If the FCC approves the petition, the channel will be added to the Table of Allotments and will become available for application. Television stations currently operate on VHF Channels 2 through 13, or UHF Channels 14 through 69, with power depending on the channels used (VHF or UHF), and station separations determined by three zones. However, after the conversion to digital television (DTV) is complete, television stations will operate only on channels 2 through 51. Channels 52 through 69 will be used for other purposes(though the possibility remains that some of this spectrum may be used for broadcasting by licensees who bid for it at auction. All new applications for television stations must be for DTV facilities.

TV Translators/Low Power TV (LPTV) Stations

There are two types of secondary television stations—*television translators* and *Low Power Television* (LPTV) stations. A TV translator is a station that rebroadcasts signals from full-service television stations. LPTV stations may rebroadcast a full-service station television signal and may also originate their own programming. LPTV stations may also operate as a subscription service. These two secondary television services may operate on any available VHF or UHF channel, provided they do not cause objectionable interference to full-service stations or other authorized translator or LPTV stations.

How to Apply for an AM, FM or TV Broadcast Station

To apply for a new AM station, or a commercial FM station or a commercial TV broadcast station for which a vacant allotment is available, the applicant must submit FCC Form 301 (Application for Construction Permit for Commercial Broadcast Station), with the appropriate application filing fee. Noncommercial educational television applicants, and noncommercial educational FM station applicants filing for operation on Channels 201 through 220, must submit FCC Form 340 (Application for Construction Permit for Noncommercial Educational Broadcast Station). No application filing fee is required for noncommercial educational applicants.

These forms are also used to make modifications in an existing facility. FM translator, TV translator, and LPTV applicants must submit FCC Form 346 (Application for Authority to Construct or Make Changes in an LPTV, TV Translator, or FM Translator Station). These forms require information about the citizenship, legal, and financial qualifications of the applicant, as well as engineering and technical specifications of the proposed or modified transmitter site.

Another way to obtain a broadcast station is to purchase an existing station that the owner is willing to sell. The FCC does not maintain a list of stations for sale and does not participate in the negotiations of the sales contract.

Station brokers and communications attorneys can assist in identifying stations that are for sale. You may also contact individual station owners directly to see if they are interested in selling their station. After you have found a station for sale, and have signed a contract to purchase the station, FCC Form 314 (Application for Consent to Assignment of Broadcast Construction Permit or License) must be submitted within 30 days accompanied by appropriate filing fees. Applicants who apply to purchase a station may not take over operation until the FCC approves the application to purchase the station.

Once the application is approved, the buyer must submit a letter of consummation within 90 days of the grant. FCC Form 323 (Ownership Report for Commercial Stations) or FCC Form 323-E (Ownership Report for Noncommercial Educational Stations) must also be submitted within 90 days of the grant. FCC Form 315 (Application for Consent to Transfer of Control of Corporation Holding Broadcast Station Construction Permit or License) must be submitted when a controlling block of shares of a broadcasting company is transferred to a new entity or an individual. There is also a FCC Form 316 (Application for Consent to Assignment or Transfer of Control) which is used when a station is involuntarily transferred, such as to a trustee in bankruptcy. FCC Form 316 is also used for Pro Forma (changes in form, not substance) assignments and transfers, such as a sale from a person to a corporation controlled by that person.

Broadcast applications must be submitted in triplicate, with the appropriate application filing fee attached. (See *Mass Media Fee Filing Guide* for fee information, fees, and filing forms. These are available from the Public Service Division or the FCC Call Center at 1-888-CALL-FCC.) Applications for noncommercial educational stations do not require a filing fee.

Applicants Must Give Local Notice

All applicants must give local notice in a newspaper of general circulation in the community in which the station is licensed or proposed to be licensed. They must also afford an opportunity for the public to file comments on these applications with the Commission. Copies of the application must be maintained in the station's public files or at a location accessible to the public in the community where the station is proposed, (i.e. public library or post office). Licensees who submit a license renewal application must give local public notice of the filing by broadcasting announcements over their stations. Any application that is designated for hearing must again give local public notice indicating such action by the Commission. (See 47 CFR Section 73.3580).

What Happens Once an Application is Filed With The FCC?

Applicants for new and major modifications in existing facilities of commercial and noncommercial AM, FM, and TV broadcast stations are tendered for filing and then placed on a cutoff list once the application is accepted for filing. The cut-off Public Notice of Acceptance triggers a 30 day filing period for any competing applicants interested in filing an opposition or petition to deny any of the applications listed on the public notice.

For the AM expanded band, the Commission issued a public notice in 1992 establishing a filing window for existing AM stations to file a petition to migrate. The Commission ranked all the petitions to migrate in accordance with the station's priority group and the interference improvement factor. Based on the overall ranking of the petitions, the Commission produced an Allotment Plan, and issued a public notice (which was later revised twice) identifying the stations receiving an allotment authorization. The Commission gave these stations until June 15, 1997, to file applications for construction permits to build facilities in the expanded band.

Applications generally are processed in the order in which they are filed. They are reviewed for engineering, legal, and financial data by the Mass Media Bureau, which under delegated authority acts on routine applications. If an application is defective, the applicant will be afforded one opportunity to amend their application. Thereafter, if all defects have not been corrected, the application will be returned. In most instances all filing fees are retained by the Commission regardless of final disposition of the application. If an application is complete and has no defects, petitions to deny, or competing applications, the application may be granted without hearing and a Construction Permit will be issued. All dispositions of applications are announced by the Commission.

Petitions for Reconsideration of grants made without hearing can be filed within 30 days of the release date of the public notice of such grant. However, when an objection is filed, the petitioner must show good cause why the objection was not raised before the grant.

Hearing Procedure

In instances where it appears that an application does not conform to Commission Rules and Regulations, and petitions to deny or competing applications are filed, the Commission may designate the application for hearing. The Commission will issue a Public Notice listing the applications that are designated for hearing. The hearing fee must be paid within 60 days of the release of that Public Notice. The hearing notice generally allows the applicant 20 days to file a "Notice of Appearance." The Administrative Law Judge (ALJ) hearing the case is also assigned at this time. Amendments to the application may be filed as a matter of right within 30 days, with a showing of good cause, if the amendment relates to issues first raised in the Hearing Designation Order. After the hearing procedure and review of evidence and statements, the ALJ issues an Initial Decision. Initial Decisions may be contested within specified time limits. In all cases heard by an ALJ, the Commission or a review board it establishes may hear oral arguments and may adopt, modify or reverse the ALJ's Initial Decision. Any action taken by the Commission may be appealed to the District of Columbia Appellate Courts.

Construction Permits

When an application is granted, either by Bureau staff or as a result of a comparative hearing, there is a specified period of time given the grantee for completing the construction of the station. Construction permits for radio are issued for 18 months, 24 months for television. Grantees must request call sign assignment when the construction permit has been granted. If no request is received from the grantee for a specific call sign, the Commission will automatically assign one. If for whatever reason a grantee of a construction permit cannot complete the construction of the station within the specified months, the grantee may request an extension of time on FCC Form 307 (Application for Extension of Broadcast Construction Permit or to Replace Expired Construction Permit) provided the grantee justifies the request showing the amount of completion and estimated amount of time needed to complete construction. Upon completion of construction, the grantee may begin program tests. However the appropriate form must be on file within 10 days of commencement of program tests. This procedure for program tests applies only to nondirectional stations. Directional stations must receive approval from the Commission prior to commencing program tests. For AM, FCC Form 302-AM (Application for New Broadcast Station License); for FM, FCC Form 302-FM (Application for FM Broadcast Station License); or for TV, FCC Form 302-TV (Application for TV Broadcast Station License).

Comparative Hearings

In April 1996, the Commission adopted an order implementing Section 204(a) of the Telecommunications Act of 1996 ("Telecom Act"). The Telecom Act adopts new Section 309(k) of the Communications Act, which eliminates comparative renewal hearings and directs the Commission to grant a broadcaster's renewal application if statutory renewal standards are met. This new procedure effects a major change in the way the Commission processes renewals. With respect to broadcast renewal applications filed after May 1, 1995, the statute eliminates comparative renewals and establishes, instead, a new twostep renewal procedure. It also codifies specific standards for the Commission to apply in considering broadcasters' renewal applications. Additionally, Section 204(a)(2) of the Telecom Act amends Section 309(d) of the Communications Act to make the standard for filing petitions to deny conform to the new statutory renewal standards. The Commission's rules will be modified to conform to the renewal procedures and standards adopted in the Telecom Act.

The Telecom Act does not define the terms contained in the renewal standards embodied in Section 309(k), and the Commission likewise does not define those terms in this order. It is the Commission's present intent to continue to apply existing policy statements and case law, refining these as appropriate on a case-by-case basis in interpreting the statutory terms that govern the new renewal process. If the Commission determines at some future time that further clarification of the renewal standards is appropriate, it will conduct such proceedings as may be warranted.

Pending comparative renewal proceedings and mutual exclusivities involving applications filed on or before May 1, 1995, will be concluded pursuant to the current rules, and accordingly, the Commission will leave intact procedural provisions of the current rules that refer to comparative renewal proceedings until those pending proceedings and exclusivities are finally resolved. The Commission also terminated BC Docket No. 81-742 in which the Commission had been considering reform of the comparative renewal process.

Licenses and Ownership

All radio and television stations are licensed for eight years. With regard to ownership, on March 8, 1996, the Commission released an order implementing Sections 202(a) and 202(b)(1) of the Telecommunications Act of 1996. These provisions direct the Commission to eliminate the national multiple radio ownership rule and modify the local radio ownership ("radio contour overlap") rule. With the order, the Commission conformed its rules to these provisions of the Telecom Act. The new rules became effective upon their publication in the *Federal Register*, which occurred on March 15, 1996.

Previously the Commission's rules limited commercial radio ownership on a nationwide basis to 20 AM stations and 20 FM stations, except that minority owners were permitted to own up to 25 AM and 25 FM stations. Nonminority owners, in addition to being permitted to own up to 20 AM and 20 FM stations, were permitted to hold an attributable interest in 5 additional stations per service, provided that those additional 5 stations were small business-controlled or minority-controlled. Section 202(a) of the Telecom Act directs the Commission to eliminate its rules limiting the number of AM or FM broadcast stations that may be commonly owned nationally and the order implemented this provision.

Previously, with regard to local ownership, the Commission's rules permitted ownership of up to three commercial radio stations, no more than two of which may have been in the same service, in radio markets with 14 or fewer stations, provided that the owned stations, if other than a single AM and FM station combination, represented less than 50 percent of the stations in the market; in markets with 15 or more commercial radio stations, ownership of up to two AM and two FM commercial radio stations was generally permitted if the combined audience share of the commonly owned stations did not exceed 25 percent in the market. Section 202(b)(1) of the Telecom Act requires the Commission to revise its rules to provide that:

- In a radio market with 45 or more commercial radio stations, a party may own, operate, or control up to 8 commercial radio stations, not more than 5 of which are in the same service (AM or FM).
- In a radio market with between 30 and 44 (inclusive) commercial radio stations, a party may own, operate, or control up to 7 commercial radio stations, not more than 4 of which are in the same service (AM or FM).
- In a radio market with between 15 and 29 (inclusive) commercial radio stations, a party may own, operate, or control up to 6 commercial radio stations, not more than 4 of which are in the same service (AM or FM).
- In a radio market with 14 or fewer commercial radio stations, a party may own, operate, or control up to 5 commercial radio stations, not more than 3 of which are in the same service (AM or FM), except that a party may not own, operate, or control more than 50 percent of the stations in such market.

Accordingly, the Commission implemented these changes. In so doing, it noted that it was not altering any of its other rules regarding the local ownership of radio stations.

On November 7, 1996, the Commission released a Notice of Proposed Rule Making in MM Dockets No. 96-222, 91-221, and 87-8, reviewing the Commission's national broadcast television ownership rules. Pursuant to the Telecommunications Act of 1996, an entity may generally not have an attributable interest in commercial television stations with an aggregate national audience reach exceeding 35 percent. A station's audience reach is determined by the number of television households in that station's Area of Dominant Influence (ADI), as determined by the commercial ratings service Arbitron. The *Notice* first addressed the "UHF discount," by which a UHF station is attributed with only 50 percent of the television households in the market, and postponed further consideration of whether to modify or repeal the rule until 1998 (at which time all ownership rules will undergo a statutorily-mandated biennial review. Second, the *Notice* proposed to eliminate the satellite exemption to the rule for satellite stations that operate in separate markets from their parent stations. However, for same-market satellites, the Commission proposed not to double count the audience in that market when calculating the parent station's national audience reach. Similarly, where one station has a local marketing agreement (LMA) with another, the *Notice* proposed not to double count that market. Finally, the Commission proposed to define a station's market for the purposes of calculating audience reach by using Designated Market Areas (as determined by Nielsen) instead of ADIs, because Arbitron no longer updates its county by county determinations of each station's ADI.

Auctions for Certain Mutually Exclusive Initial License Applications

The Balanced Budget Act of 1997 provides, with respect to certain Mutually Exclusive (MX) applications for an initial license or construction permit filed on or after July 1, 1997, that the Commission *must* grant the license or permit to a qualified applicant through a system of competitive bidding. The Balanced Budget Act specifies, however, that the Commission's competitive bidding authority does not extend to licenses or construction permits for certain public safety radio services, for noncommercial educational broadcast and public broadcast stations, or for digital television service given to existing terrestrial broadcast licensees to replace their analog television service licenses.

With respect to such applications that were filed before July 1, 1997, the Balanced Budget Act (1) *authorizes* the Commission to resolve mutually exclusive applications for an initial construction permit for a commercial radio or television broadcast station through competitive bidding, rather than through the comparative hearing process; (2) specifies that the persons filing such applications shall be treated as the only persons eligible to be qualified bidders; and (3) directs the Commission to waive any FCC rules necessary to permit such persons to enter into an agreement to procure the removal of a conflict between their applications during the 180-day period beginning on the day of the Balanced Budget Act's enactment.

RULES AND REGULATIONS

FCC Rules may be obtained by writing or calling the Government Printing Office, Washington, DC 20402; (202) 512-1800. The Rules on FCC Practice and Procedures are contained in the Code of Federal Regulations (CFR), Title 47, Parts 0–19. Broadcast rules are found in Title 47 of the CFR Parts 70–79. An on-line version of the CFR rules is available at the U.S. House of Representatives Internet Law Library Home Page. To search in the CFR, (e.g. find rule section 47 CFR Section 73.1226), enter the search item as 47:cite w/ 3 73.1226. To search for a whole part, (e.g. Part 73), enter the search item as 47:cite w/3 73. To use wildcard, (e.g. to find all 73.500 rule sections), enter the search item as 47:cite w/3 73.5*. Note that if a whole part is to be retrieved, the maximum number of records retrieved will need to be increased (e.g., to 400 for Part 73). Items will not be retrieved in order by CFR section number. Please be aware that FCC rules received from this site may not incorporate recent revisions.

Application Forms

Application forms may be obtained by calling the automated forms request line at 1-800-418-3676 or 1-202-418-3676 (recording machine available 24 hours a day, seven days a week), writing to Federal Communications Commission, Forms Distribution Center, 9300 E. Hampton Drive, Capital Heights, MD 20743, or using the FCC's Internet on-line Forms Request link at http://www.fcc.gov/formpage.html. Questions concerning filing procedures should be directed to the Public Service Division, Consumer Assistance Branch, 1919 M Street, N.W., Washington, DC 20554; telephone: (202) 418-0200. Alternatively, questions may be referred to the Audio Services Division for the radio services [(202) 418-2720 for technical questions, (202) 418-2710 for legal questions] or the Video Services Division at (202) 418-1600.

THE RULE MAKING PROCESS

The FCC, like all other federal government agencies, enacts new rules and regulations through the terms of the Administrative Procedures Act (APA). The APA specifies how rules may be proposed, adopted, and appealed. The APA assures that the public has input into the rule making process. Part I of the FCC Rules and Regulations provides detailed information on the FCC's general rules of practice and procedure.

Initiation of Actions

Rule making actions can be proposed by anyone through a *Petition for Rule Making*. Such petitions bring the desires of individuals or groups to the attention of the Commission. The FCC will evaluate the petitions and either dismiss them for one of the reasons in Section 1.401 of the Rules or accept them for action. If accepted, a *Public Notice* will be released giving a brief description of the details of the petition. Other rulemakings may begin by direction of Congress, the President, or the courts. The FCC may also initiate rule makings on its own motion.

Early Options

Initiation actions are handled in one of four ways. A petition can be dismissed directly through a *Memorandum Opinion and Order* (MO&O). Minor rule changes of an administrative nature can be made effective directly through a *Report and Order* (R&O). Generally, however, either a *Notice of Inquiry* (NOI) or a *Notice of Proposed Rule Making* (NPRM) will be drafted by the staff. If enough information is at hand to draft the proposed rule changes, an NPRM will be issued without first gathering background material in an NOI. After adoption of an NPRM by vote of the Commissioners, the *ex parte* rules become effective. Because this inhibits direct contact between FCC staff and concerned parties, the NOI may be considered an essential first step in many matters. *Ex parte* does not apply at the NOI stage. Comment periods are provided to allow interested parties to express their views. If, based on a review of the comments received, the Commission finds that an NOI should be terminated without further action, it will issue an MO&O stating the reasons for the termination. Otherwise, the matter will be considered for specific rule making in an NPRM. The MO&O is subject to a 30 day period in which parties may file petitions for reconsideration if they have good reason to believe the Commission acted incorrectly.

Notice of Proposed Rule Making (NPRM)

Most rule making items begin as an NPRM. The staff prepares the NPRM for consideration by the five Commissioners. An NPRM normally presents the issues and alternatives, may ask specific questions to help finalize the matter, and sets forth proposed rules or rule amendments. An NPRM may address one or multiple topics. Comments must be filed by the close of the commend period, usually 30 days. Reply comments may then be filed in response to comments, usually for a 15 day period. These two comment periods provide the Commission with a "written debate" of the issues. The comments and reply comments received are reviewed and enter heavily into the Commission's final actions. However, a comment does not represent a vote for the proposed rules. The Commission must decide on each issue based on the public"s interest, convenience, and necessity. Even if the majority of the comments oppose an item, the FCC can nevertheless adopt the proposal. The FCC can also schedule items for oral argument among the interested parties.

Filings should be clear, concise, and address the issues in the NPRM. Emotional responses or uninformed responses are of little value. Copies of the actual texts of the NPRM need to be studied before making comments. All NPRMs are published in the *Federal Register* and are available at most larger libraries or law libraries.

The comments are reviewed by Commission staff at the Branch or Division levels. Recommendations are then passed to the Bureau level and ultimately to the Commissioners. The Commissioners will then approve, disapprove, or direct the staff to modify the recommendations. The staff or the Commissioners may also decide that the record still does not support a decision, and a *Further Notice of Proposed Rule Making* (FNPRM) will be released. Like an NPRM, the comment periods apply. Once the record supports a final decision, the FCC will act by the adoption of a *Report and Order.*

Report and Order (R&O)

An R&O may be used to dismiss the proposals in the NPRM, adopt some of the proposals (possibly modified based on the comments), or adopt the rules as proposed. Rules will normally not become effective until 30 days after publication of the R&O in the *Federal Register.* The R&O constitutes the Commission's final action unless a petition for reconsideration is filed.

Petition for Reconsideration

Any interested person may petition for reconsideration of a final action. The petitions for reconsideration must be filled within 30 days of the date the R&O was published in the *Federal Register.* The Commission will act on a petition for reconsideration by *Memorandum Opinion and Order.* The initial decision may be modified or the petition may be denied in the MO&O.

Ex Parte Considerations

The FCC revised its *ex parte* rules June 2, 1997. The *ex parte* rules govern the manner in which persons may communicate with the Commission concerning the issues in a proceeding. The rules play an important role in protecting the fairness of the FCC's proceedings by assuring that FCC decisions are not influenced by impermissible off-the-record-communications between decision-makers and others. At the same time, the rules are designed to ensure that the FCC has sufficient flexibility to obtain the information that is necessary for it to make reasonable decisions. The *ex parte* rules apply to anyone who engages in the kind of communications covered by the rules whether or not they are a party to the proceeding. The rules govern *ex parte* presentations to FCC decision makers. The rules are found in sections 1.1200 to 1.1216 of Title 47 of the Code of Federal Regulations, as amended by FCC 97-92, released March 19, 1997.

In some types of proceedings (*restricted proceedings*), the rules prohibit *ex parte* presentations to decision makers concerning the issues in its proceedings. In other types of proceedings (*permit-but-disclose proceedings*), the rules require that summaries of such presentations be placed in the record. In some types of proceedings (*exempt proceedings*), there are no restrictions on *ex parte* presentations. The rules describe which types of restrictions or requirements, if any, apply to *ex parte* presentations in the various types of Commission proceedings.

A presentation is a communication directed to the merits or outcome of a proceeding, including any procedural or other issues raised in the proceeding—there are some exceptions, however. Communications which are inadvertently or casually made are not presentations. Neither are routine inquiries about compliance with the Commission's procedural rules, such as when a pleading must be filed, so long as the question has not become the subject of dispute in the proceeding.

Inquiries relating solely to the status of a proceeding are not presentations. However, a status inquiry is deemed a presentation if it states or implies a view as to the merits or outcome of the proceeding or a preference for a particular party, states why timing is important to a particular party (other than the need to avoid administrative delay), or indicates a view as to the date by which a proceeding should be resolved.

An *ex parte* presentation is any presentation which, if written (including electronic mail), is not served on the parties to the proceeding, or, if oral, is made without advance notice to the parties and without opportunity for them to be present.

Decision-making personnel are those people at the Commission who are or who may reasonably be expected to be involved in formulating a decision, rule, or order in a proceeding. All FCC Bureau or Office staff are considered decision-making personnel with respect to decisions, rules, and orders in which their Bureau or Office participates unless they have been designated as part of a separate trial staff or otherwise formally excluded from the decisional process in the proceeding.

Certain kinds of *ex parte* presentations are considered "exempt" and may be freely made even if a proceeding is permit-but-disclose or restricted. Participants in Commission proceedings should be aware that these include:

- Presentations authorized by statute or by the Commission's rules to be made without service, as well as the filing of required forms
- Presentations to the General Counsel's office concerning judicial review of a proceeding
- Presentations directly relating to emergency situations
- Situations where confidentiality is necessary to protect persons making *ex parte* presentations from possible reprisals

Participants in Commission proceedings are also permitted to make presentations requested by (or made with the advance approval of) the Commission or staff for the clarification or presentation of evidence, or for resolution of issues, including possible settlement, subject to certain disclosure requirements. Parties to restricted Commission proceedings are permitted to orally request that the Commission act by a particular date or give reasons for expediting the case other than the need to avoid administrative delay. If they do so, however, they must promptly file a detailed summary of the presentation in the record and serve it on the other parties to the proceeding. Other parties may respond similarly in support or opposition to the request. The rules contain other exemptions and should also be consulted to determine if a particular exemption contains a disclosure requirement.

Ex parte presentations are permissible in "exempt" proceedings. There are six kinds of exempt proceedings:

- Notice of Inquiry proceedings.
- Petitions for Rulemaking (except for petitions requesting the allotment of a broadcast channel) and other requests for rule changes and policy statements.
- Tariff proceedings (including directly-associated waiver requests or requests for special permission) prior to being set for investigation.
- Proceedings relating to prescription of common carrier depreciation rates under section 220(b) of the Communications Act prior to release of a public notice of specific proposed depreciation rates.
- Informal complaint proceedings under section 208 of the Communications Act.
- Complaints against a cable operator regarding its rates that are not filed on FCC Form 329.

Ex parte presentations are permitted in "permit-but-disclose" proceedings provided that the presentations are disclosed in the manner specified in the rules. There are 11 kinds of permit-but-disclose proceedings:

- Informal notice and comment rulemaking proceedings other than a proceeding for the allotment of a broadcast channel, upon release of a Notice of Proposed Rulemaking.
- Proceedings involving a rule change, policy statement, or interpretive rule adopted without a Notice of Proposed Rule Making upon release of the order adopting the rule change, policy statement, or interpretive rule.
- Declaratory ruling proceedings.
- Tariff proceedings which have been set for investigation under section 204 or 205 of the Communications Act (including directly-associated waiver requests or requests for special permission).
- Unless designated for hearing, proceedings under section 214(a) of the Communications Act that do not also involve applications for authorizations for radio communications services.
- Unless designated for hearing, proceedings involving applications for Cable Landing Act licenses that do not also involve applications for authorizations for radio communications services.
- Proceedings involving a request for information filed pursuant to the Freedom of Information Act.
- Proceedings before a Joint Board or proceedings before the Commission involving a recommendation from a Joint Board.
- Proceedings conducted pursuant to section 220(b) of the Communications Act for prescription of common carrier depreciation rates upon release of a public notice of specific proposed depreciation rates.
- Proceedings to prescribe a rate of return for common carriers under section 205 of the Communications Act.
- Cable rate complaint proceedings pursuant to section 623(c) of the Communications Act where the complaint is filed on FCC Form 329.

If there is only one party involved in a permit-but-disclose proceeding, that party may communicate with the Commission without disclosure. If, however, a presentation is made by a non-party or by one of multiple parties in such a proceeding, the presentation is deemed *ex parte* and must be disclosed.

There are specific procedures for ensuring that the presentations are reflected in the record of a permit-but-disclose proceeding. *Ex parte* presentations by members of Congress or their staffs and other federal agencies or their staffs need be disclosed only if they are of substantial significance and clearly intended to affect the ultimate decision in the proceeding. Disclosure of *ex parte* presentations by members of Congress or their staffs and other federal agencies or their staffs will generally be made by the Commission's staff.

In other circumstances, persons making written *ex parte* presentations must, no later than the next business day after the presentation, submit two copies of the presentation to the Commission's secretary under separate cover for inclusion in the public record. The presentation (and cover letter) must clearly identify the proceeding to which it relates, (including the docket number, if any) indicate that two copies have been submitted to the Secretary; and be labeled as an *ex parte* presentation. If the presentation relates to more than one proceeding, two copies must be filed for each proceeding.

Persons making oral *ex parte* presentations must disclose them if they present data or arguments not already reflected in that person's written comments, memoranda or other filings in that proceeding. In that case, the person must, no later than the next business day after the presentation, submit to the Commission's Secretary, with copies to the Commissioners or Commission employees involved in the oral presentation, an original and one copy of a memorandum which summarizes the new data or arguments. The subject matter of the presentation must be fully disclosed; a mere listing of the subjects discussed is not sufficient, and more than a one or two sentence description of the views and arguments presented is required.

The memorandum (and cover letter) must clearly identify the proceeding to which it relates (including the docket number, if any), indicate that an original and one copy have been submitted to the Secretary, and be labeled as an *ex parte* presentation. If the presentation relates to more than one proceeding, two copies of the memorandum (or an original and one copy) must be filed for each proceeding. There is an exception to these requirements where, for example, presentations occur in the form of discussion at a widely attended meeting, and preparation of a memorandum as specified in the rule might be cumbersome. Under these circumstances, the rule may be satisfied by submitting a transcript or tape recording of the discussion as an alternative to a memorandum.

The Commission's Secretary issues a Public Notice at least twice a week listing any written *ex parte* presentations or written summaries of oral *ex parte* presentations in permit-but-disclose proceedings. It is possible that some presentations might inadvertently be omitted from this list, so interested persons should review the public file or record in proceedings about which they are concerned, where the copies of written presentations and memoranda of oral presentations can be found.

In all proceedings other than those specifically listed as "exempt proceedings" or "permit-but-disclose proceedings," *ex parte* presentations are prohibited. In such proceedings, called "restricted proceedings," any written presentations to FCC decision-makers must be served on all the parties and oral presentations may not be made unless all parties are given advance notice and an opportunity to be present. If there is only one party involved in a restricted proceeding, that party may freely communicate with the Commission, because there are no other parties entitled to advance notice or service.

However, presentations by a non-party or by one of multiple parties in a proceeding are prohibited unless all parties are served or given an opportunity to be present. Restricted proceedings include all proceedings that have been designated for hearing, proceedings involving amendments to the broadcast table of allotments, applications for authorizations for radio communications services, and all waiver proceedings (except for those directly associated with tariff filings).

Certain circumstances make you a party with a right mto service of written presentations and advance notice of oral presentations in restricted proceedings and to disclosure of presentations in permit-but-disclose proceedings. Being a party for purposes of the ex parte rules does not necessarily mean that you will be considered a party for other purposes. Parties include people who file applications, petitions, or other requests for relief and those making responsive filings. To make someone a party, such an application or petition must be directed to the Secretary, the relevant Bureau or Office, or the Commission as a whole—not just to the Chairman or individual Commissioners. In addition, to make someone a party, responsive filings must reference the original filing and be served on the filer. For example, persons who file petitions to deny license applications are parties. In addition, persons who file "informal objections" under rule section 73.3587 are parties if their objections are served. (However, the rules contain a special exception for individual listeners or viewers who file informal comments about pending broadcast applications. Such persons are not made parties even if their informal comments are served, and they do not have to serve their comments on the applicant.)

When a Freedom of Information Act request is filed, the person who filed the request is a party and, where the requested information is the subject of a request for confidentiality, the person filing the request for confidentiality is also a party. Applicants and people filing mutually-exclusive applications are parties with respect to each others' applications except in services that the Commission has announced will be subject to competitive bidding or lotteries.

People who file complaints and the subject of the complaint are parties if the complaint is a formal complaint under § 208 of the Communications Act or the complaint shows that it was served on the subject of the complaint. (Earlier, it was noted certain kinds of complaints are treated as exempt.) A person who files a petition to revoke a license or other authorization or who files a petition for an order to show cause and the subject of the petition are parties. The subject of an order to show cause, hearing designation order, notice of apparent liability, or similar notice or order, or petition for such notice or order is a party.

In informal notice and comment rulemaking proceedings (except for those involving the allotment of a broadcast channel), and proceedings before a Joint Board or before the Commission to consider the recommendation of a Joint Board, members of the general public after the issuance of a notice of proposed rulemaking or other order are parties. In rulemakings for the allotment of a broadcast channel, the petitioners and those making responsive filings are parties.

The sunshine period prohibition imposes a restriction on presentations to Commission decision-makers in addition to the limitations otherwise applicable under the *ex parte* rules. It applies to proceedings in which the Commission has issued a Public Notice that a matter will be considered at a Commission meeting—that is, the matter has been placed on the "Sunshine Agenda." While the sunshine period prohibition is in effect, all presentations to decision-makers concerning matters listed on a Sunshine Agenda, whether *ex parte* or not, are prohibited unless they fall within certain exceptions. In particular, presentations are permitted if they occur in the course of a widely attended speech or panel discussion and concern a Commission action in an exempt or a permit-but-disclose proceeding that has been adopted. Members of Congress or their staffs and other federal agencies or their staffs may make presentations in exempt or permit-but-disclose proceedings, but presentations that are of substantial significance and clearly intended to affect the ultimate decision in the proceeding must be disclosed, either by the presenter or the Commission's staff.

The sunshine period prohibition applies from the release of a Public Notice that a matter has been placed on the Sunshine Agenda until the Commission releases the text of a decision or order relating to the matter, issues a Public Notice stating that the matter has been deleted from the Sunshine Agenda, or issues a Public Notice stating that the matter has been returned to the staff for further consideration, whichever occurs first.

Appeal to the Courts

After the Commission has considered and reconsidered a matter, interested parties may appeal the decision to the federal courts. Ultimately, the Supreme Court could hear the case. Under current law, however, the Commission need not wait for the court decision before enacting the new rules and regulations. As long as the matter has been given full consideration under the Administrative Procedures Act, the FCC may place the new rule in effect until a court rules to the contrary or orders a "stay."

1.4
FCC COMPLIANCE AND INFORMATION BUREAU

BY DAVID E. WILSON
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC

OVERVIEW

The Compliance and Information Bureau (CIB) of the Federal Communications Commission (FCC) is responsible for providing information to the general public, resolving telecommunications interference problems and enforcing the FCC's technical rules. It also operates the network of FCC field offices throughout the United States.

Since the eighth edition of this handbook, many changes have occurred to the FCC's field office structure.

The most significant of these are:

The closing of numerous field offices across the country and their replacement with resident agents.

The elimination of individual public contact points at each field office and their replacement with a centralized National Call Center in Gettysburg, Pennsylvania.

The bureau name changed from Field Operations Bureau to the Compliance and Information Bureau.

In the course of this restructuring the number of CIB employees nationwide was reduced from nearly 400 to approximately 250.

As a result of these changes, the country is now divided into three regions of CIB responsibility—the Western Region, the South Central Region, and the Northeast Region. The Western Regional Director is responsible for the field offices in Denver, Los Angeles, San Diego, San Francisco, and Seattle, in addition to the resident agents in Anchorage, Honolulu, and Portland. The South Central Regional Director is responsible for the offices in Atlanta, Dallas, Kansas City, New Orleans, and Tampa, as well as the resident agents in Houston, Miami, Norfolk, and San Juan. The Northeast Regional Director is responsible for the offices in Boston, Chicago, Columbia (Maryland), Detroit, New York, and Philadelphia, and for the resident agents in Buffalo and St. Paul. Each of the cities covered by resident agents has two agents located there. There are no FCC offices in these cities, instead, all work is conducted out of the agents' homes.

Geographic Area Covered by Each Field Office

Table 1.4-1 provides an alphabetical listing of all states and territories under FCC jurisdiction. The table shows which FCC field office is responsible for each of the counties in every state and territory.

The mailing addresses for each of these offices are listed below. All telephone calls to the CIB should be directed to the FCC's National Call Center in Gettysburg (1-888-CALL-FCC).

Atlanta Office
Federal Communications Commission
3575 Koger Boulevard, Suite 320
Duluth, GA 30096-4958

Boston Office
Federal Communications Commission
1 Batterymarch Park
Quincy, MA 02169-7495

Chicago Office
Federal Communications Commission
Park Ridge Office Center, Room 306
1550 Northwest Highway
Park Ridge, IL 60068-1460

Columbia Office
Federal Communications Commission
9200 Farm House Lane
Columbia, MD 21046

Dallas Office
Federal Communications Commission
9330 LBJ Freeway, Room 1170
Dallas, TX 75243-3429

Denver Office
Federal Communications Commission
165 South Union Boulevard, Room 860
Lakewood, CO 80228-2213

Detroit Office
Federal Communications Commission
24897 Hathaway Street
Farmington Hills, MI 48335-1552

Kansas City Office
Federal Communications Commission
8800 East 63rd Street, Room 320
Kansas City, MO 64133-4895

Los Angeles Office
Federal Communications Commission
Cerritos Corporate Tower
18000 Studebaker Road, Room 660
Cerritos, CA 90701-3684

New Orleans Office
Federal Communications Commission
2424 Edenborn Avenue, Suite 460
Metarie, LA 70001

New York Office
Federal Communications Commission
201 Varick Street, Suite 1151
New York, NY 10014-4870

Philadelphia Office
Federal Communications Commission
One Oxford Valley Office Building, Room 404
2300 East Lincoln Highway
Langhorne, PA 19047-1859

San Diego Office
Federal Communications Commission
Interstate Office Park
4542 Ruffner Street, Room 370
San Diego, CA 92111-2216

San Francisco Office
Federal Communications Commission
3777 Depot Road, Room 420
Hayward, CA 94545-2756

Seattle Office
Federal Communications Commission
11410 NE 122nd Way, Room 312
Kirkland, WA 98034-6927

Tampa Office
Federal Communications Commission
2203 North Lois Avenue, Room 1215
Tampa, FL 33607-2356

Table 1.4-1
FCC Field Office Responsible for Each State/Territory

State	Area Served
Alabama	All counties are served by the Atlanta field office.
Alaska	All counties are served by the Seattle field office.
American Samoa	All counties are served by the San Francisco field office.
Arizona	All counties are served by the San Diego field office.
Arkansas	All counties are served by the New Orleans field office.
California	Kern, Los Angeles, Orange, San Bernardino San Luis Obispo, Santa Barbara, and Ventura counties are served by the Los Angeles field office.
	Imperial, Riverside, and San Diego counties are served by the San Diego field office.
	Alameda, Alpine, Amador, Butte, Calaveras, Colusa, Contra Costa, Del Norte, El Dorado, Fresno, Glenn, Humboldt, Inyo, Kings, Lake, Lassen, Madera, Marin, Mariposa, Mendocino, Merced, Modoc, Mono, Monterey, Napa, Nevada, Placer, Plumas, Sacramento, San Benito, San Francisco, San Joaquin, San Mateo, Santa Clara, Santa Cruz, Shasta, Sierra, Siskiyou, Solano, Sonoma, Stanislaus, Sutter, Tehama, Trinity, Tulare, Tuolumne, Yolo, and Yuba counties are served by the San Francisco field office.

Table 1.4-1 (*Continued*)
FCC Field Office Responsible for Each State/Territory

State	Area Served
Colorado	All counties are served by the Denver field office.
Connecticut	All counties are served by the Boston field office.
Delaware	Kent, Sussex, and New Castle (below C&D Canal) counties are served by the Columbia, Maryland, field office.
	New Castle (above C&D Canal) county is served by the Philadelphia field office.
District Of Columbia	All of the District is served by the Columbia, Maryland, field office.
Florida	Escambia and Santa Rosa counties are served by the Atlanta field office.
	All other counties are served by the Tampa field office.
Georgia	All counties are served by the Atlanta field office.
Guam	All counties are served by the San Francisco field office.
Hawaii	All counties are served by the San Francisco field office.
Idaho	All counties are served by the Seattle field office.
Illinois	All counties are served by the Chicago field office.
Indiana	Allen, De Kalb, Elkhart, Fulton, Kosciusko, La Grange, Marshall, Noble, St. Joseph, Steuben, and Whitley counties are served by the Detroit field office.
	All other counties are served by the Chicago field office.
Iowa	All counties are served by the Kansas City field office.
Kansas	All counties are served by the Kansas City field office.
Kentucky	Bath, Bell, Boone, Bourbon, Boyd, Bracken, Breathitt, Campbell, Carter, Clark, Clay, Elliott, Estill, Fayette, Fleming, Floyd, Franklin, Gallatin, Garrard, Grant, Greenup, Harlan, Harrison, Jackson, Jessamine, Johnson, Kenton, Knox, Knott, Larel, Lawrence, Lee, Leslie, Letcher, Lewis, Lincoln, Madison, Magoffim, Martin, Mason, McCreary, Menifee, Montgomery, Morgan, Nicholas, Owen, Ownsley, Pendleton, Perry, Pike, Powell, Pulaski, Roberston, Rockcastle, Rowan, Scott, Wayne, Whitley, Wolfe, and Woodford counties are served by the Detroit field office.
	All other counties are served by the Chicago field office.
Louisiana	All counties are served by the New Orleans field office.
Maine	All counties are served by the Boston field office.
Mariana Islands	All counties are served by the San Francisco field office.
Maryland	All counties are served by the Columbia, Maryland, field office.
Massachusetts	All counties are served by the Boston field office.
Michigan	Alger, Baraga, Delta, Dickinson, Gogebic, Houghton, Iron, Keweenaw, Marquette, Menominee, Ontonagon, and Schoolcraft counties are served by the Chicago field office.
	All other counties are served by the Detroit field office.
Midway Islands	All counties are served by the San Francisco field office.
Minnesota	All counties are served by the Chicago field office.
Mississippi	All counties are served by the New Orleans field office.
Missouri	All counties are served by the Kansas City field office.

Table 1.4-1 (*Continued*)
FCC Field Office Responsible for Each State/Territory

State	Area Served
Montana	All counties are served by the Seattle field office.
Nebraska	All counties are served by the Kansas City field office.
Nevada	All counties are served by the San Francisco field office.
New Hampshire	All counties are served by the Boston field office.
New Jersey	Bergen, Essex, Hudson, Hunterdon, Mercer, Middlesex, Monmouth, Morris, Passaic, Somerset, Sussex, Union, and Warren counties are served by the New York field office. Atlantic, Burlington, Camden, Cape May, Cumberland, Gloucester, Ocean, and Salem counties are served by the Philadelphia field office.
New Mexico	All counties are served by the Denver field office.
New York	All counties are served by the New York field office.
North Carolina	All counties are served by the Kansas City field office.
North Dakota	All counties are served by the Chicago field office.
Oklahoma	All counties are served by the Dallas field office.
Ohio	All counties are served by the Detroit field office.
Oregon	All counties are served by the Seattle field office.
Pacific Trust Territories & Commonwealth	All counties are served by the San Francisco field office.
Pennsylvania	All counties are served by the Philadelphia field office.
Puerto Rico	All counties are served by the Tampa field office.
Rhode Island	All counties are served by the Boston field office.
South Carolina	All counties are served by the Atlanta field office.
South Dakota	All counties are served by the Denver field office.
Swains Island	All counties are served by the San Francisco field office.
Tennessee	All counties are served by the Atlanta field office.
Texas	All counties are served by the Dallas field office.
Utah	Emery, Garfield, Grand, Kane, Piute, San Juan, Sevier, and Wayne counties are served by the San Diego field office. All other counties are served by the San Francisco field office.
Vermont	All counties are served by the Boston field office.
Virgin Islands	All counties are served by the Tampa field office.
Virginia	Arlington, Fairfax, Loudoun, and Prince William counties are served by the Columbia, Maryland, field office. All other counties are served by the Kansas City field office.
Wake Island	All counties are served by the San Francisco field office.
Washington	All counties are served by the Seattle field office.
West Virginia	All counties are served by the Columbia, Maryland field office.
Wisconsin	All counties are served by the Chicago field office.
Wyoming	All counties are served by the Denver field office.

FCC National Call Center

All telephone inquiries to the CIB field offices are now handled through the FCC's toll-free National Call Center in Gettysburg, Pennsylvania. The number for the call center is: 1-888-CALL-FCC (1-888-225-5322). Broadcasters may call this number to check the status of pending license applications. They may also use it to report interference problems and request FCC enforcement action. For example, recently the CIB received complaints from many broadcasters about pirate radio operations. These complaints led to FCC enforcement actions across the country.

COMMON TECHNICAL COMPLAINTS ABOUT BROADCASTERS

The complaint-resolution aspect of the CIB's function has a flip side to it—CIB also handles complaints received from the public about broadcasters. Usually the complaints that come from the public regarding broadcasters deal with non-technical issues such as program content. However, some public complaints are related to the technical side of broadcasting. For example, complaints about broadcast signals causing telephone interference and blanketing interference are very common. To maintain good public relations, and avoid FCC penalties for non-compliance, broadcasters should be particularly aware of their responsibilities in these two common complaint areas.

Blanketing interference is a phenomenon that occurs when a strong signal from a nearby transmitter overloads the front end of a receiver and thus prevents the receiver from receiving other signals. The FCC rules clearly define the extent of the area around a radio broadcast antenna that is considered to be blanketed. For AM stations, this area is the region encompassed by the 1 V/m contour (see Section 73.88). For FM stations, it is the 562 mV/m contour (see Section 73.318). No specific contour has been defined as the *blanketing contour* for TV stations, although TV broadcasters are required to correct reasonable complaints of blanketing interference within the immediate vicinity of new or recently modified transmission facilities (see Section 73.685(d)). Broadcasters who build new or modified facilities that require an FCC construction permit are required to correct all blanketing interference problems for which complaints are received within one year after program tests begin.

Broadcasters are not required to correct interference to telephones unless it is the rare case of a receiver in a cordless telephone being "blanketed" by a nearby broadcast transmitter. Neither standard wired telephones nor cellular telephones are covered by the FCC's blanketing rules. Nevertheless, in order to maintain a good relationship with their community, many broadcasters do provide assistance to consumers who experience interference to their home phones.

As this *Handbook* goes to press, the FCC is considering changes to its blanketing rules in MM Docket 96-62. Broadcasters frequently should consult a current version of the FCC's rules and regulations to determine if their blanketing responsibilities have changed.

The *Interference to Home Electronic Entertainment Equipment Handbook*, Bulletin CIB-2 is a publication that helps explain interference issues to the general public. The booklet describes various types of interference to consumer electronic devices and recommends various corrective actions. It also provides information about suppliers of RF filtering devices. It is available from the FCC's website at www.fcc.gov and the U.S. Government Printing Office.

SELF-INSPECTION CHECKLISTS

The CIB role that is probably of most concern for many broadcasters is its program of inspecting broadcast facilities. Most CIB inspections are conducted without prior notice to the licensee. Most times these inspections will involve a visit to the station's studios, transmitter and/or transmitter control points. However, on some occasions, the FCC may simply monitor the station's signal off the air for compliance with modulation and spurious emission limits.

The CIB has prepared self-inspection checklists for AM, FM and TV stations. These checklists provide descriptions of the most frequently-violated broadcast regulations, references to the rule sections where these regulations can be found, and check-boxes that station personnel may use to keep track of their compliance status for each item. Copies of these checklists can be obtained from any CIB field office or from the FCC's website.

FCC FINES

In 1997, the FCC revised its guidelines for assessing forfeitures in cases of non-compliance with its rules. A base forfeiture amount is assigned for various types of violations. However, the forfeiture amount can be adjusted upward for one or more of the following reasons:

Egregious misconduct
Ability to pay/relative disincentive
Intentional violation
Substantial harm
Prior violations of any FCC requirements
Substantial economic gain
Repeated or continuous violation

The fine can also be adjusted downward for these reasons:

Minor violation
Good faith or voluntary disclosure
History of overall compliance
Inability to pay

The types of violations most likely to concern broadcasters, and their associated base forfeiture amounts, are listed in Table 1.4-2. These amounts are for each violation. The reader should note, however, that in cases where a particular violation continues for multiple days the FCC considers each day of the continuing violation to be a separate violation. Also, the FCC has the option of applying alternative or additional sanctions as permitted by law.

Table 1.4-2
Base Fine Amounts for Various FCC Violations

Violation	Fine
Misrepresentation/lack of candor	$25,000
Construction and/or operation without an instrument of authorization for the service	$10,000
Failure to comply with prescribed lighting and/or marking	$10,000
Violation of public file rules	$10,000
Violation of political rules: reasonable access, lowest unit charge, equal opportunity, and discrimination	$9,000
Unauthorized substantial transfer of control	$8,000
Violation of children's television commercialization or programming requirements	$8,000
EAS equipment not installed or operational	$8,000
Failure to maintain directional pattern within prescribed parameters	$7,000
Violation of main studio rule	$7,000
Violation of broadcast hoax rule	$7,000
AM tower fencing	$7,000
Failure to permit inspection	$7,000
Transmission of indecent/obscene materials	$7,000
Interference	$7,000
Exceeding of authorized antenna height	$5,000
Fraud by wire, radio or television	$5,000
Unauthorized discontinuance of service	$5,000
Use of unauthorized equipment	$5,000
Broadcasting telephone conversations without authorization	$4,000
Exceeding power limits	$4,000
Failure to respond to Commission communications	$4,000
Violation of sponsorship ID requirements	$4,000
Unauthorized emissions	$4,000
Using unauthorized frequency	$4,000
Failure to engage in required frequency coordination	$4,000
Construction or operation at unauthorized location	$4,000
Violation of requirements pertaining to broadcasting of lotteries or contests	$4,000
Violation of transmitter control and metering requirements	$3,000
Failure to file required forms or information	$3,000
Failure to make required measurements or conduct required monitoring	$2,000
Violation of enhanced underwriting requirements	$2,000
Failure to provide station ID	$1,000
Unauthorized pro forma transfer of control	$1,000
Failure to maintain required records	$1,000

ALTERNATIVE INSPECTION PROGRAM

The FCC has teamed up with some state broadcaster associations to implement a self-inspection program that enables broadcasters to avoid surprise FCC inspections. The Alternative Inspection Program (AIP) is a voluntary program that provides broadcasters with an opportunity to demonstrate compliance with the FCC rules without an actual FCC inspection and the accompanying threat of citations and fines. Under the AIP, a state broadcaster association-designated person, who is not an FCC employee, will inspect a station at the station's request. There is usually a nominal fee associated with the inspection. If the broadcaster is found to be in compliance with the FCC rules, a Certificate of Compliance will be issued and the FCC will be notified. If the station is found not to be in compliance it will be notified of the reason(s) for its failure but the FCC will not be notified of the inspection. The FCC field offices have agreed not to conduct a surprise inspection of any station that has received a Certificate of Compliance under the AIP for two or three years after the certificate is received. The exact duration of this "no surprise inspection" period depends on the particular field office involved. A field office might still inspect a station that has received a certificate, however, if it has a reason to believe that the station is not operating in compliance, such as when the FCC receives a complaint about the station's operation.

Broadcasters who wish to participate in an AIP should contact their state broadcaster association to find out if the program is available in their state.

1.5
FREQUENCY COORDINATION

ROBERT VAN BUHLER
KNIX-FM/TEMPE, ARIZONA

RICHARD RUDMAN
RADIO STATION KFWB, LOS ANGELES, CALIFORNIA

WHAT IS FREQUENCY COORDINATION?

The push to leave the studio to originate programs has been a part of broadcasting from the beginning. Now with over 12,000 radio and 1,500 TV stations on the air, the demand for reliable ways to get news, sports, public affairs, and entertainment programming from the field back to the studio has naturally increased. Part 74 of the FCC's Rules deals with segments of the spectrum that have been assigned to the Broadcast Auxiliary Service (BAS) to accommodate this demand. Part 74 also covers other services outside the scope of this discussion, such as FM and TV translators, Low Power TV, ITFS, and experimental broadcast stations.

While the number of stations requiring support spectrum has grown tremendously since 1970, the amount of spectrum allocated for this support has decreased. Broadcasters using Part 74 spectrum began to experience competition for the limited number of channels allocated not only among broadcasters, but also among the other users who share some of the Part 74 spectrum such as cable operators. Interference from other broadcasters has gradually become a critical problem in many markets and there are no provisions for exclusive licensing in Part 74. Consequently, frequency coordination is necessary to facilitate the allocation of limited spectrum for a growing number of uses among a growing number of users.

The coming age of personal communication brings new pressures. The FCC has already designated a portion of broadcast auxiliary 2 GHz spectrum for other users. While there may be some hope for replacement spectrum, there are no longer any certainties. We must face the fact that we may not have sufficient spectrum to meet our own growing needs. The reality of digital television as well as the prospect of digital AM and FM mean that our needs for auxiliary spectrum for high quality transmission will certainly grow. The need to keep our coordination efforts alive and active will be greater than ever ass pressures from both sides increase. We should also look carefully at the very technologies that are driving spectrum refarming for solutions to our needs.

Some History of Broadcast Coordination

Formal coordination has existed for some time for major events such as the Olympics, political conventions and space launches. In the case of the political conventions, the major national television networks have rotated the chairmanship of an ad hoc group, known as the National Political Conventions Frequency Coordinating Committee. Washington, D.C. has had unique coordination problems along with a history of cooperative effort for pool feeds to cover events of national interest. Such committee activity has traditionally arisen out of a need, mutually felt by broadcast engineers, to discuss concerns before they become problems. Among the oldest groups formed to deal with mutual operational problems, including coordination, were the Washington Executive Broadcast Engineers (WEBE) and the TV Broadcasters All-Industry Committee in New York.

The first major region to form a committee to deal exclusively with Part 74 coordination was Southern California. In 1976, the FCC revised Part 74, Subpart D, of its Rules, which deals with the Remote Pickup Service used by radio and television stations for dispatch, on-air remote pickup, cues and orders, communications and transmitter telemetry. Recognizing the use of narrowband FM telemetry. Recognizing the use of narrowband FM technology, the Commission reallocated the UHF (450 and 455 MHz) frequencies from wideband 100 kHz channels to smaller bandwidth 50 kHz, 25 kHz and 10 kHz channels. Under the FCC's 1976 changes, R and S channels were intended for program material, and cues and orders necessary to implement that programming. N1 and N2 channels were to be used for broadcast program material, cues and orders necessary to that material, and operational communications. Microwave path setup and dispatch functions fall under this last category. P channels may be used only for Operational Communications and telemetry. In 1980 the FCC granted the Southern California Frequency Coordinating Committee (SCFCC) a waiver which further split the S and N1 channel as shown in Table 1.5-1.

Note that since the SCFCC plan allows for center

Table 1.5-1
1976 FCC 450 Channel Plan Showing SCFCC Waiver Changes

Channel Waiver Designation	Channel Bandwidth	Peak FM Deviation	Total # of CHS.	Total # of SCFCC Channels
N1	50 kHZ	10 kHz	12	0
N2	25	5	24	60
R	50	10	10	14
S	100	35	2	0
P	10	1.5	8	8

channel use of the split N1 50 kHz channels, the maximum number of 25 kHz channels can jump from 24 to 60 since the center channel can be used when geographic separation and or terrain shielding permit.

The FCC will now accept license applications based on channel segments of 5 kHz. These 5 kHz segments can be stacked for the Part 74 VHF high band and for UHF band to form channels of the required bandwidth for a given use. Users should consult either their local coordination group or the FCC's Auxiliary Branch in Washington about a local band plan if questions arise. Regions considering formation of a coordination group can study existing band plans based on segmented channels before implementing new plans. No examples of band plans are given. Plans must be custom tailored for each region.

Coordination activity spans many aspects of broadcasting. Some regions have not experienced congestion and interference with their own licensees, but have organized committees to deal with problems associated with broadcasters who enter their regions for temporary operation. After conducting a survey at the chapter level on how this problem is viewed locally, the Society of Broadcast Engineers (SBE) has for many years identified frequency coordination as a chapter project. The SBE National Frequency Coordinating Committee (NFCC) was formed to assist any chapters interested in this project. In addition, the SBE NFCC publishes quarterly a list of volunteer coordinators who are responsible for over 100 regions in 50 states, plus Puerto Rico.

If there is one adjective that can be used to describe the nature of effective frequency broadcast coordination that word is local. Electronic news gathering (ENG) for both radio and television introduces many unknowns. Mobile paths cannot be predicted from day to day, or even hour to hour. Large scale news, public affairs, or sporting events may mean interference for everyone involved unless technical personnel familiar with a number of factors are consulted. Since many of these factors are dependent on local propagation anomalies, terrain considerations, and user patterns, experience has proven that engineers who work with these factors at the local level on a day-to-day basis are the best people to consult with and to ask for advice for successful operations.

Propagation Characteristics

A number of factors can modify the expected line-of-sight characteristics of VHF and UHF aural and visual transmission. Many of these can be understood best if the radio waves as affected are compared to light travelling through the reflective surfaces of a periscope so the viewer can see around corners.

The three most common propagation anomalies are: (1) ducting effects, (2) reflection due to thermal inversion layers and (3) obstacle gain (sometimes referred to as knife edge refraction).

These anomalies can occur for seconds at a time, or may exist in a predictable manner for much longer periods. For instance, obstacle gain can be used in some cases to provide permanent and quite dependable links. In the Part 74 environment, these anomalies have to be recognized and taken into account to assure co-channel and adjacent channel users of adequate protection for both fixed and mobile links. Known anomalies in a region must be factored into any coordinated environment. As in the case of terrain, licensees should seek out engineers who have longtime experience in the region to coordinated so anomalies that have been identified, and the paths they affect, can be taken into consideration.

THE BROADCAST COORDINATION PROCESS

Successful coordination efforts are all founded on licensee-to-licensee contact. The coordination effort, however it is structured, acts to facilitate this contact. As of this writing, Part 74 of the FCC's Rules places the burden of preventing interference squarely on the licensee. Since broadcasting is highly competitive, no local committee should place itself in a position where it assigns spectrum. Eligible radio, TV, cable and other selected entities have an equal right to use shared frequencies specified under Part 74 or part 78 of FCC Rules and Regulations, subject only to the service and classifications contained therein. No action taken by a coordination committee should restrain these privileges in any way.

Unless otherwise determined by future FCC actions, committees are voluntary facilitators of the coordination process. The committee has no authority to force adoption of its recommendations. Successful committees never act as policemen. This role is left to the FCC if all other means of settling disputes fails. However, committees can and do use neutral mediators to try to solve the inevitable problems and disputes that arise in a competitive world. The committee exists to facilitate coordination, not to enforce it.

A chaotic situation would already exist in most markets if the engineering community had not developed fundamental coordination techniques and practices. Many cable and network entities, production originally users, and others who now have access to spectrum intended to serve only the needs of local broadcasters would not be able to operate reliably otherwise.

The broadcast or cable engineer does not operate in a purely technical environment when it comes to frequency coordination. It is an art and technical discipline that extends beyond the skills normally found in the engineering department. There are other key players whose cooperation and participation is essential to efficient frequency coordination.

A news director can create a climate of efficient spectrum management in a radio or TV operation by stressing the importance of adhering to good operating practice and home channel plans.

Both news and engineering managers must realize that frequency (or channel) information must be shared by users for coordination to succeed. The ultimate goal is interference-free operation for all users. This goal must be kept in mind whenever Broadcast Auxiliary Spectrum/Cable TV Relay Service (BAS/CARS) frequencies are employed. Cooperation from all parties, and an active and supportive leadership example from engineering and news managers in charge, is essential. Radio/TV station and cable general managers can make frequency coordination possible by allowing engineers the time and financial support needed to get the job done. Many progressive local broadcast associations have seen value in underwriting the costs of frequency coordination by providing computer equipment, postage, stationery, duplication expenses and other support to make frequency coordination a success.

General managers are vital to the process in another way. In a growing number of stations, engineering is provided by a contractor whose services are limited to maintenance and repair. By nature of their contract, these providers have very little input on the design of mobile systems or fixed links. In such cases, it is important that the manager insist that his BAS/CARS equipment applications and operation be handled harmoniously with the needs of the local coordinating committee and the broadcast and cable community it serves.

Station or system managers should know if their engineers are involved in frequency coordination. The best approach is to meet with the engineer and discuss the importance of coordination to the well being of the broadcast and cable community.

Simply requiring chief engineers to attend regular coordination meetings is helpful to the coordination process. By attending the meetings, the engineer is familiarized with the process and becomes acquainted with license changes and coordination problems occurring in the market.

In addition to the reasons mentioned that might lead broadcasters to start coordination activity, one more must be added. Many broadcasters now travel outside their markets to cover news, sports, or public affairs events. In 1983 the FCC revised Section 74.24 of the Rules to make it easier for broadcasters to cover such events. A footnote to this Rule change asked that regional coordinators make themselves known to the Commission. The SBE NFCC was first formed to compile a list of coordinators to go beyond the Commission's request.

COORDINATOR CONSIDERATIONS FOR VARIOUS BROADCAST USES OF PART 74 SPECTRUM

Dispatch and Aural Remote Pickup

Channels for this purpose are currently available in a range from 26 MHz to and including 455 MHz.

In major markets experiencing congestion, and in other markets where older equipment is still in service, 26 MHz operation for dispatch and even some types of remote pickup are still being done. Lower power cueing systems for TV field operations sometimes appear in this band. Since it is near to present Citizen's Band channels, it is subject to interference from this service. Propagation conditions at 26 MHz include skip so it is common to experience interference in this band from hundreds, or thousands, or miles away at certain times of the year.

VHF dispatch and aural remote pickup are more common throughout the country. With the advent of excellent commercial narrowband FM mobile transceivers and remotely operated base stations on hills and mountains, many radio and TV stations have used Part 74 VHF channels reliably for years.

Committees will be confronted with a number of seemingly conflicting uses and users as they sort through the years of accumulated problems in VHF and UHF bands. One guiding principle that has helped many committees is the order of priorities outlined by the FCC in Section 74.403(b). The transmission of remote pickup program material for broadcast, either live or delayed, takes precedence over other permissible uses.

VHF and UHF remote pickup with usable audio frequency response to 15 kHz is available with a variety of equipment. In many regions, interference from adjacent channels licensed to other services makes such operation difficult. A strong adjacent channel signal can capture a FM receiver, pulling the discriminator away from the desired signal. As channel use builds in a region, broadband white noise rises, making life difficult for wideband receivers.

Some equipment marketed for broadcast use was not designed for today's levels of channel congestion and sometimes makes the problem appear worse than it really is.

Transmitters should always be connected to antennas through ferrite isolation devices and harmonic filters to minimize generation of mixing products in solid state finals. This applies to repeaters, base stations, and mobile transmitters. This was not as important in the days of tube finals, but is critical in today's nonlinear solid state world. Ferrite isolators are used for simplex transmitters. Ferrite circulators are used for repeaters.

Receivers can use combinations of attenuators, helical tuned circuits and cavity filter to protect against first stage overload from out of channel emissions. This is especially important since the 450 MHz Part 74 allocations are within 1 MHz of Land Mobile dispatch and paging bands. Paging transmitters operating at 250 W commonly employ two-way sites also used by broadcasters and can easily overload an unprotected receiver.

Broadband white noise at common sites is now being viewed as much of a culprit to degraded receiver performance as spurious emissions. As the number of transmitters at a site increases, the noise level at the site increases. This noise desensitizes all receivers at the site. If a user is operating a receiver at a common site used for high power paging transmitters that are on the air most of the time, the receiver may be trying to listen in a channel less than 1 MHz away from a 250 W transmitter with a high gain antenna. The resultant 1000 W (ERP) can cause problems for even the best equipment. Committees can educate members on this subject as well.

At quiet sites, receivers with GAsFET preamplifiers working with preselector/filter network receivers have been used successfully to increase receiver sensitivity or regain insertion losses from various filter elements. Careful design can prevent overloading the receiver with the add-on preamplifer.

Remotely located (satellite) receivers will give users better coverage and better immunity to various types of interference. It is desirable to locate satellite receivers at receive-only locations. Such locations can be selected after spectrum analysis shows that adjacent channel and out-of-band emissions are well within the receiver's ability to function at an optimal performance level. Receive-only locations are also good places to utilize high gain GAsFET preamplifiers. Local coordination groups may be of service in locating such sites, and maintaining their RF integrity. Getting the receiver closer to the transmitter is always preferable to increasing transmitter power as a technique to improve reliability. Leased program circuits can deliver audio back to the studio.

Committees can educate members on the pitfalls of common sites, and work with site managers to let them know of licensee concerns. There is a trend in many markets toward re-engineering common sites using transmitter and receiver combiners, and physical separation of transmit and receive antennas. These changes can improve the transmit and receive performance at most sites. Changes like this are expensive. A committee can educate users on the benefits of such projects, and help generate support to make them happen.

Committees can also make members aware of non-Part 74 licensing options for broadcasters who require two-way communications which are not related to Part 74 priorities such as the business band spectrum or cellular telephones. Cellular telephones have been used by many stations for some years now for coverage of news and even program length material. The FCC makes no distinction between cellular and wired telephones for the purpose of broadcast transmissions as of this writing. Stations should check, however, to see how and if broadcast use of cellular phones conforms with state telephone tariffs. Extended cellular transmissions from any site that can lock up a number of cells may be an abuse of the cellular system.

The two blocks of UHF channels available to broadcast licensees are separated by 5 MHz. This is adequate for repeater operations, and many licensees in markets with hilly or mountainous terrain have employed this type of system to allow field units to communicate with the base no matter where they are. The convention of repeater pair frequencies is that the 455 group be used for the repeater output. Many markets have taken great pains to correct reversed situations for mutual benefit. Simplex channels have to treated on a case-by-case basis. Repeater operation has the benefit of being truly wireless, since remote base systems usually depend on telephone company local channels for control and audio. Unfortunately, repeater operation requires two channels.

Spectrum efficiency in the UHF bands by splitting channels is an accepted mode of operation in some markets. Since it is possible to drive audio response beyond 5 kHz from a 25 kHz narrowband FM channel, many radio stations which want better audio quality for news reports can get it without using a 50 kHz R channel, provided the receiver used is well into full quieting. Committees can work with their members to make sure everyone is using the least amount of bandwidth consistent with needs.

Similarly, a station engaged in voice dispatch does not need a 50 kHz N1 channel. A 25 kHz N2 channel is more than adequate for this use. Responsible committees can help assure that use is matched to bandwidth.

Narrow band radio systems that exhibit better that 5 kHz audio response with 5 kHz FM deviation made possible use of splinter or offset channels where distance separations and relative transmitter power levels would permit. High performance receiver IF filters have been in use for this purpose since the early '80s, usually as a part of audio processing modifications to FCC type accepted transmitters that required an FCC modified type acceptance. Built on a firm base of existing rugged mobile equipment, such systems proved that further channel splitting was practical. Receivers designed to take advantage of the spectrum efficiencies of low FM modulation indexes are now coming to market. These receivers promise even greater quality improvements along with greater spectrum efficiency.

High Band VHF and UHF Coordination

Committees can sometimes arrange sharing of channels between users. This can even involve several licensees operating a community repeater system. This choice not only helps enhance spectrum efficiency, but may mean small stations can become part of a larger system they could not otherwise afford. The following are guidelines for new users of Part 74 communications:

1. The potential user should understand that he must be a part 73 licensee, or a network (as defined in Part 74) to qualify for Part 74 channels.
2. Know that the committee really exists to help accommodate new users in bands already crowded—not to keep new users out.
3. The potential user should explain his needs during an open committee meeting, and publish what is said in the newsletter so everyone concerned can be aware of what is going on. It may be that someone at the meeting might make a suggestion to the potential user that could improve his system, or even accommodate his needs by some other means. For instance, cellular mobile telephone might be all someone needs to do ENG microwave setup if there are no dispatch channels left in the market.

4. The committee may recommend certain frequencies. At this point, it is up to the potential licensee to contact each co-channel and adjacent channel occupant and complete the transaction. It is NOT the responsibility of the committee to do this.
5. In the event the parties reach an impasse regarding an issue, the committee can and should act as a mediating force. Usually a compromise can be arranged.

In a congested RF environment, there is always the possibility of adjacent channel interference. Mobiles operating in the field can find themselves next to other mobiles transmitting on an adjacent channel. If this channel happens to be on the repeater output of the first station, the person in the mobile cannot communicate with its base while the other mobile has its transmitter on. If this is a chronic problem, one station may find it desirable to change frequencies. No amount of after the fact coordination can ever solve all of these problems.

There is at least one operating practice a committee can promote that will help: make it a standard in the market for each licensee to use the appropriate station identification and unit designator call sign. Not only is this in compliance with FCC Rules, but call signs make it easier to track down interference problems that will inevitably arise.

Wireless Microphones

Wireless microphone are becoming more common and should not be overlooked in the coordination process. Coordinators are often asked to act as a clearinghouse for major events where a large number of such mikes are in use. Wireless microphone operation is a permitted use in the band specified in Part 74.802.

Non-RF Alternatives

There are non-RF alternatives available for STL systems. Laser links are feasible for very short distances where there is little likelihood of any interruption of the line-of-sight path.

Many stations still rely on program channels leased from their local telephone company. For short distances that do not involve very many telephone exchanges, this may still represent the best solution for some stations. This may be true especially in markets that have reached STL saturation.

TV MICROWAVE CONSIDERATIONS

Television has come to rely on their own or common carrier microwave links to interconnect their studios with their transmitters and to facilitate live remote broadcasts. Rules governing links operated by broadcasters are found in Part 74 of the FCC Regulations and cover frequencies from 2, 2.5, 5, 6, 4, 7, 13, 18, 21, 23 and 40 GHz bands.

Microwave transmissions can accommodate one or more subcarriers imbedded in the video signal to carry mono or stereo television sound, aural service for a coo-owned AM or FM station, program cueing, telemetry, or a combination of these elements. In congested markets one should use the lowest possible subcarrier frequency combinations to keep occupied bandwidth as low as possible.

In major markets the number of stations engaged in field operations exceeds the number of available channels. While spectrum is available in the 2, 2.5, 6.4, 7, 13, 18, 23, and 40 GHz bands at this time, many stations and coordinating committees in these major markets have opted to make 2 GHz the primary band for ENG. However, this band will have some level of STL and Inter-City Relay (ICR) activity for the foreseeable future due to its long haul propagation characteristics.

The use of 2 GHz for ENG with its better propagation characteristics means stations have to install fewer ENG receive locations in their service area. Path lengths up to 50 miles can be operated reliably and under optimum conditions, including low humidity, path lengths up to 100 miles may be possible.

In many markets, there is intense competition for the seven primary 2 GHz channels (1.9 through 2.1 GHz) and for the 2 (+1 grandfathered) channels at 2.5 GHz. The 2.5 GHz band is shared with other user such as the industrial, scientific, and medical (ISM), service and the private operational fixed microwave service.

Stations having 2 GHz STLs have been, for the most part, relocating to 7 or 13 GHz for several reasons. First, they may themselves be involved in 2 GHz ENG. Second, a fixed link in this band has a high potential risk of interference from mobile operations. Since modern TV ENG portable equipment is frequency agile, and non-technical personnel are more involved in field operations, this is a very real risk.

Coordination for ENG is being done more and more on a real time basis. Coverage of breaking news cannot be planned. Coupled with the limited number of 2 GHz channels, coordination will only be more difficult in the future.

Suggested Operational Guidelines for TV ENG Operations—The Home Channel Plan

1. Form a Microwave Subcommittee. Compile a list of persons responsible for ENG technical and operational decisions at each station. This list should contain telephone numbers providing access 24 hours a day. This list should be circulated to all stations involved in ENG operations and be updated as needed.
2. The Microwave Subcommittee should be a forum to devise a Home Channel Plan for the region. The purpose of this plan is to give each user a primary channel for his operation. Alternate polarization is assigned to each channel. Split channel operation should be the long range goal for the market. Every user is furnished with a copy of this plan for reference.

3. As part of this plan, each user should set a goal to become frequency, polarization and power agile.
4. Once the Home Channel Plan has been developed a method of coordinating this plan in real time is needed. This can be accomplished by a hard wired ring down network between ENG control points, by using a two-way radio system with base stations located at the control point of each ENG license or by establishing and maintaining a list of phone numbers for ENG control points of all participating stations.
5. Each licensee should equip the ENG operation at the following level:
   - Mobile transmitters should be frequency agile.
   - Each transmitter shall have continuously variable control of output power.
   - Each mobile truck should ideally be equipped with an antenna with excellent sidelobe suppression of the silhouette type.
   - Each mobile truck should be capable of changing the polarity of their feed.
   - Each truck should have a means of identifying its test bars, and ideally be able to identify its transmission in the vertical interval. If this latter type of equipment is employed, the ID is viewed by unlocking the vertical hold control of a video monitor.
   - Use frequency agile ENG receive sites with remote steerable antennas.
   - Employ multiple receive sites.
   - Use steep-skirted channel filters ahead of receivers.
6. Truck crews are instructed never to light up until they receive permission from their control point, even on their home channel.
7. Truck crews should receive careful instruction on polarity settings; using only enough power for the path involved, and in the overall operational agreements with other stations engaged in ENG activity.
8. Stations normally stay on their home channel. They use the ENG control point communication system to request different channels for extra activity, to eliminate inevitable adjacent channel interference during some operations, or for paths that conflict on their home with ICR links, or STLs. The request is made of the station which is assigned the desired channel as its Home Channel.
9. Remember, no one really owns a channel. Anyone might be called upon to share his or her channel on a real time coordination basis.

The plan must enjoy unequivocal support from all levels of management. It is imperative that this support be articulated and understood throughout the entire ENG operation, both in technical and non-technical areas, verbally and in writing.

All reasonable requests for temporary sharing of channels must be accommodated.

Spectrum use for all major events should be preplanned. When advance notice of a major event is given, a coordinator is assigned. Requests for channel assignment are solicited, operating channel assignments are made, and contingencies considered and addressed.

Examples of this would include community celebrations of a regional or national interest, a papal visit to a particular market or the opening of the United Nations General Assembly. On a larger scale, examples would include Olympic coverage or political conventions. The foregoing are documented examples of successful coordination.

Temporary Operation for TV Microwave RPU

As in the aural services, temporary operation away from a station"s service area without a special license is permitted for up to 720 hours or 30 days per year under Section 74.24 of the rules. As mentioned earlier in this chapter, the Coordinator Listing compiled and maintained by the SBE can make coordination for this type of operation less painful. The coordination burden is still on the licensee who is visiting not to cause interference to the local stations.

During this type of operation, visitors should use their broadcast on-air call sign, not their Part 74 call sign. They should work closely with the local coordination group. Confirm all communications in writing.

A visiting broadcaster should also obtain an accurate database to determine existing co-channel and adjacent channel users. If such a database is not available through a local coordinator, it is incumbent on the temporary user to accurately ascertain who would be affected.

It is important to notify existing co-channel and adjacent channel users of the intended link giving transmitter power output, dates and times of operation, and antenna type and gain information. Exchange names, addresses and 24-hour telephone numbers.

An engineering study of the path should be performed. This should include detailed path calculations. Several computer programs using accepted path consideration algorithms are available from a number of sources, including the SBE NFCC.

A field spectrum analysis is a valuable tool for engineering any link, fixed or temporary. It may uncover licensees who have not been entered in the database, newly authorized construction that may interfere with the path, temporary users who have not notified affected licensees or the local coordination group, Section 74.24 users, or multipath from co-channel links on different paths.

This rule is primarily designed to cover special events and should not be substituted for filing licenses.

Major Event Frequency Coordination for RF Equipment

The FCC has traditionally recognized ad-hoc coordination groups for many years for events such as the national political conventions, major sporting events such as the Olympic Games and the America's Cup races, and news events like NASA shuttle launches and landings.

Such events require a great deal of planning and hard work. Preparation for the national conventions typically begin 24 to 18 months in advance, although no substantive work can commence until the city sites are announced.

The FCC's normal course of action for such events is to suspend Section 74.24 of the Rules for a period before and during the event. This has the effect of requiring all Part 73 licensees wanting to use RF equipment of any type at the event to contact and coordinate with other users via the ad-hoc coordination group.

SBE, through its Washington, D.C. attorney will assist local groups with making the proper request and showing to the FCC for such a waiver to be issued for an event.

If an event takes place in an enclosed venue, it may be possible to use channel in use by local broadcasters outside the venue. Using very low power transmitters is an excellent way to allow a number of users to coexist inside such a venue. A large number of hand-held transceivers can often be used in adjacent channels by simple expedient of removing the rubber duck antenna and substituting a 50 Ω termination resistor. TV microwave transmitters should be cut back or attenuated to 50-11 mW. If all users cooperate, receivers in the various bands being used will be quite happy, and the needs of all users will be met with little or no interference.

Weekend events may make it possible to use TV and radio ENG channels only used during weekdays. Compatible sharing should take place in the spirit of good cooperation between engineers.

Assuring all equipment used for such events is working properly is an important consideration. While this is the responsibility of the licensee, the coordinating body can sometimes help them deal with this burden. Various techniques have proven to be effective. For example, the FCC usually sends one of its sophisticated trucks to the major political conventions to look for spurious and unauthorized emissions. NASA and the United States Air Force usually check all equipment as it comes into Edwards Air Force Base for Shuttle landings. NASA requires everyone to participate in a full light up test two hours before each landing. For smaller events, local groups commonly borrow or rent a spectrum analyzer as a service to guest broadcasters.

If it is not possible to accommodate all users within the existing Part 74 bands for major events, the FCC has granted waivers in the past for one or more unused TV UHF channels, usually for two-way dispatch. This is not possible in all markets. When this spectrum is finally allocated, this option will no longer be available.

Whatever the event, the primary goals of special event coordination are:

1. Protect the existing base of local users.
2. Accommodate itinerant users to the greatest extent possible.
3. Be present at the event to help unsnarl the inevitable glitches that will happen.
4. Report to the industry and FCC after the event so all can learn for the future.

The Benefits of Coordination for Microwave Licensees

As in the case of all Part 74 coordination, committees exist to help new and existing licensees share very limited amounts of spectrum. The move to frequency agility has made real time coordination possible for mobile operations. It mandates cooperation and sharing. In return, it allows licensees to operate their field ENG equipment and STLs with reduced risk of interference.

The Risks and Penalties of Not Coordinating

Failure to coordinate can be an expensive proposition for those who do not participate in this voluntary process. Uncoordinated mobile and fixed links have been taken off the air and re-engineered when they interfered with licensees operating under rules and procedures they agreed to abide by along with other broadcasters in the market. Precedent has been set for FCC support of local agreements where consensus is evident.

Expensive recrystalling, reprogramming, and retuning, antenna and site changes may be the only way to correct interference problems, most of which could have been avoided by communication with the coordinating committee and other users prior to the installation.

Because the broadcasting and cable industries are not monolithically structured, frequency coordination techniques cannot be identical in each market. Only general suggestions can be given as to what each manager or news director can do to participate in and support the coordination process. However, it is clear that coordination cannot succeed without the active support of general managers and news directors at whatever level is practical in a given situation.

INTERFERENCE AND THE PART 74 LICENSEE

Interference to a Part 74 licensee's operations may come from almost any source of RF one cares to mention. Radiation from unshielded computer cables has interfered with wireless microphone operation, for instance. Channel interference from other broadcasters can be thought of as a problem that can be solved by additional attempts at coordination, followed, if necessary, by mediation by a neutral party.

Locating the source of spurious emissions can be a real challenge. The committee can offer some advice and support when a suspected case of interference from an out-of-channel source is reported. Quite often, other stations in the market can make spectrum analyzers available to eliminate the possibility of a mixing product that is actually generated in the front end of the receiver due to an overload condition. This type of problem is common for both voice and video equipment.

Older solid state receivers are sometimes quite prone to front end overload. When an overload prone receiver is operating at a common site with an appreciable amount of adjacent channel activity, the receiver may be rendered useless. Filtering and attenuator pads can sometimes be suggested as cures. Video microwave receivers operating with preamplifiers can overload easily as a result of proper adjacent channel operation. In such cases, a suggestion to provide a bypass switch is in order.

Committees should encourage licensees to use call letters and unit identifiers for all voice transmissions. All video ENG trucks should be able to transmit bars with ID, and ideally should be able to place ID in the vertical interval as well as for identification during program transmission.

Tracking unidentified interference is not unlike detective work. For video, tuning through the on-air signals may provide a clue. If bars with no ID are seen at an ENG receive location, triagulation may be possible if the interference can be seen by two other locations equipped with steerable dishes. Trucks have been identified by reading street signs that appear in shots. Talent on other stations can sometimes be a good means of identification.

Interference to aural links can be challenging to trace. If call letters are not being used, anything heard then becomes a clue. If a street address is mentioned a newsroom criss-cross telephone directory can be used to find a telephone number at or near the address. A call to the number identifying yourself, your reason for calling, and if a taxicab or delivery truck has been there (or will be there) often works wonders.

Land mobile transmitters are sometimes designed so the final has power applied to it at all times. Any instability in the final can start it oscillating. These transmitters can sometimes migrate through many megahertz of spectrum, almost at random. When their exciters are keyed, they return to their proper channel. Sometimes only a spectrum analyzer and a lot of patience can trace this type of problem.

Inexpensive direction finding equipment is available for VHF and UHF tracking. Some units work with any hand-held portable. For unmodulated carriers and other types of interference that are present for long periods of time, direction finding can be a very effective (though time consuming and expensive) process.

If the licensee(s) involved, even with committee guidance, are unable to locate the interference, a call to the FCC may be in order. You will be asked for all pertinent information, such as time the interference occurs, its nature, duration, and any other details you might have learned. The FCC generally will not take interference complaints unless you can supply this information. Remember that coordinating committees and their members are never to act as police in such matters.

Local FCC offices have suffered deep staffing cutbacks that have further restricted their ability to help us find sources of willful and accidental interference. There is hope that passive tracking technology will be available in the near future. Passive tracking uses receivers located at multiple remote sites. These receivers are monitored at one location. Their signals are processed real time by computer. A source of interference can be located rapidly and accurately. The offender's RF envelope can be recorded and positively identified by its unique RF signature.

Curing Interference

Dealing with the broad topic of interference cures is beyond the scope of this chapter. When committees are working to solve such problems, creativity is definitely an asset. For instance, in a case where an aural STL on an adjacent channel on a parallel path was causing interference, the cure that was tried and worked involved installing a large dish at the transmitter. The dish was aimed slightly off azimuth from its receiver so that the receiver site being interfered with appeared in a deep null of the interfering transmit antenna.

MORE INFORMATION

For more information on frequency coordination, including a subscription to the SBE National Frequency Coordinators' Listing, or information on starting a local coordinating committee, contact the Society of Broadcast Engineers, 8445 Keystone Crossing, Suite 140, Indianapolis, IN 46240, (317) 253-1640. Website: www.sbe.org.

1.6
DISTANCE AND BEARING CALCULATIONS

DANE E. ERICKSEN, P.E., CSTRE
HAMMETT & EDISON, INC., CONSULTING ENGINEERS, SAN FRANCISCO, CALIFORNIA

INTRODUCTION

The current FCC Rules for distance calculations use two methods: *flat-earth* and *spherical-earth*. The flat-earth method assumes the distance between two points to be the hypotenuse of a right triangle whose sides are determined by the difference in latitude and longitude of the starting and ending points, multiplied by the length per degree of latitude and longitude at the mid-latitude of the two points, as shown in Figure 1.6-1. The *flat-earth* term is not meant to be disparaging, and merely refers to the use of a right triangle to calculate the distance. Because the lengths of a degree of longitude and a degree of latitude used in the flat-earth method are derived from an ellipsoid rather than a spheroid model of the earth, the flat-earth method is actually more accurate than the spherical-earth method for short to moderate distances.

The spherical-earth method uses conventional spherical trigonometry to determine the distance. Section 73.208 of the FCC Rules now requires that the flat-earth method be used for distances up to and including 475 km. Distances greater than 475 km must be calculated using the spherical-earth method, which becomes more accurate than the flat-earth method for large distances. Section 73.208 is silent on how azimuths are to be calculated, it does not specify the earth radius to be used for spherical-earth calculations.

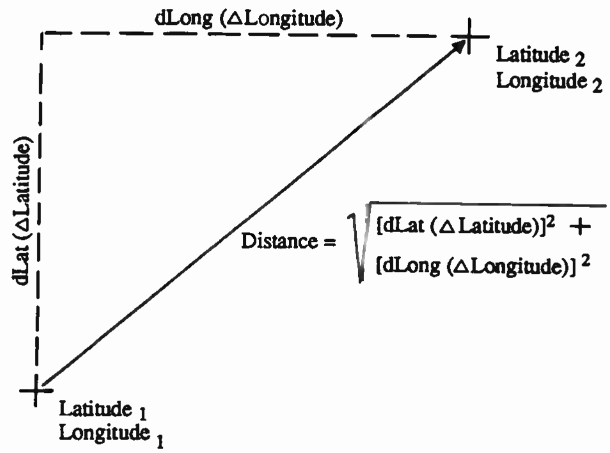


Figure 1.6-1. Flat-earth method.

FCC FLAT-EARTH METHOD

In FCC Docket 80-90,[1] formulas were substituted for the tables previously used for determining the length of a degree of latitude or of longitude as a function of latitude. However, the coefficients adopted in Docket 80-90 truncated to only two terms the trigonometric series used to generate the tables and "adjusted" the coefficients by a factor of (1.609/1.609347) because of the Docket 80-90 decision to define the conversion factor from U.S. statute miles to kilometers as 1.609,[2] rather than the value of (5,280 ft/mile) × (1200/3937 m/ft) × (1/1000 km/m), or 1.609347219 km/mile (approximately[3]).

In the *Second Report and Order* to Docket 86-144,[4] the FCC corrected these problems by adopting the full-precision, non-truncated trigonometric series for the arc length formulas given in the 1966 edition of U.S. Naval Hydrographic Office Publication Number 9 (H.O. 9) or the *American Practical Navigator* or simply *Bowditch* after Nathaniel Bowditch (1773–1838), its original author. These trigonometric series are based upon a binomial theorem expansion[5] of an ellipsoid model of the earth corresponding to the Clarke spheroid of 1866, upon which topographic maps in the United States are currently based.[6]

The trigonometric series defining the length of one degree of latitude and one degree of longitude for the Clarke spheroid of 1866 are:

$$\text{dLat} = 111.13209 - 0.56605\cos(2L) + 0.00120\cos(4L)\ldots \quad [1.6.1]$$

$$\text{dLong} = 111.41513\cos(L) - 0.09455\cos(3L) + 0.00012\cos(5L)\ldots \quad [1.6.2]$$

where dLat is the length in kilometers of one degree of latitude at latitude L and dLong is the length in kilometers of one degree of longitude, again at latitude L.

The latitude, L, is taken as the mid-latitude of the two points between which the distance is to be calculated, as follows:

$$L = (\text{Latitude}_1 + \text{Latitude}_2)/2 \quad [1.6.3]$$

where Latitude_1 and Latitude_2 are the latitudes of the starting and ending points. Similarly, Longitude_1 and Longitude_2 are the longitudes of the starting and ending points. In all cases, north latitudes are treated as positive and south latitudes as negative, and west longitudes are treated as positive and east longitudes as negative.

The distance between two points is then given by the Pythagorean theorem:

$$D = \sqrt{[(dLat)(Lat_1 - Lat_2)]^2 + [(dLong)(Long_1 - Long_2)]^2} \quad [1.6.4]$$

Plots showing how the lengths of one degree of latitude and longitude vary with latitude are given in Figures 1.6-2 and 1.6-3.

Canadian Method

In August 1987,[7] the Canadian government adopted the truncated and adjusted arc-length formulas which had been implemented by the FCC in Docket 80-90. Namely,

$$dLat = 111.108 - 0.566\cos(2L) \quad [1.6.5]$$

$$dLong = 111.391\cos(L) - 0.095\cos(3L) \quad [1.6.6]$$

Unfortunately, the Canadian Government has not yet adopted the corrected, more accurate formulas that were implemented by the FCC in Docket 86-144. Nor was the flat-earth versus spherical-earth break point changed from 350 km to 475 km. This is the source of current discrepancies between the U.S. and Canadian distance calculation methods. Although the differences between the two methods will usually not be significant when the calculated distance is rounded to the nearest kilometer (for FM) or to the nearest one-tenth kilometer (for TV), one should always check to see whether there is a difference between roundings. For calculations involving Canadian stations, the Canadian (Docket 80-90) version of the Clarke spheroid formulas is controlling.

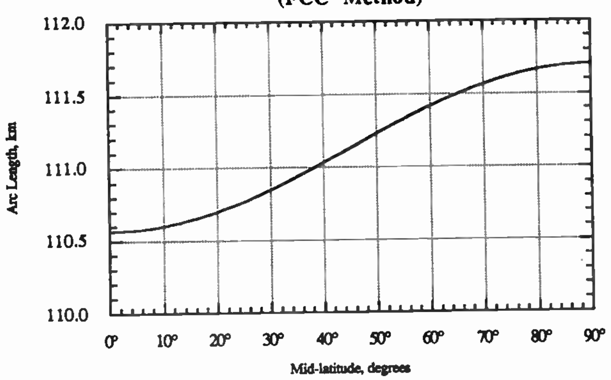


Figure 1.6-2. Length of 1° arc at latitude (FCC method).

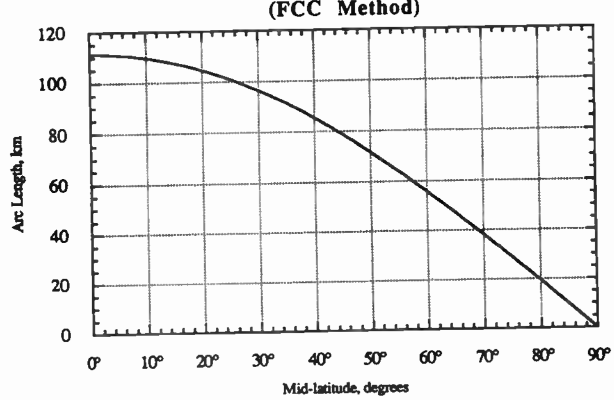


Figure 1.6-3. Length of 1° arc of longitude (FCC method).

Mexican Method

The August 11, 1992, U.S.-Mexican FM Agreement specifies the spherical-earth method exclusively; there is no provision for using the FCC flat-earth method for short to moderate distances. However, the Mexican Agreement specifies a spherical earth arc of 111.18 kilometers per degree, corresponding to an earth radius of 6,370.14 kilometers. The Mexican Agreement further specifies that azimuths are to be calculated on a spherical-earth basis. Finally, the Mexican Agreement specifies the rounding of distances to the nearest kilometer, so in that respect it matches the FCC rounding practice for FM station distances.

Spherical-Earth Method

The formula for the spherical-earth distance, or great-circle distance, is:

$$D = K\cos^{-1}[(\sin Lat_1)(\sin Lat_2) + (\cos Lat_1)(\cos Lat_2)\cos(Long_2 - Long_1)] \quad [1.6.7]$$

The constant K is in km/degree and is determined by the radius of the sphere being modeled. The FCC has never defined the earth radius to be used for spherical-earth calculations. The example given in Section 73.185(d) of the FCC Rules suggests an earth radius of 6,365 km (K = 111.090 km/degree). A 6,373 km radius (K = 111.230 km/degree) is implied by the 5,280-mile 4/3-earth radius given in Section 73.684(c)(1) of the FCC Rules. This 4/3-earth radius was also used in FCC Report No. R-6410, "Elevation and Depression Angle Tables," September 15, 1964. An earth radius of 6,367 km (K = 111.125 km/degree) can be deduced from the 1,852-meter definition of a nautical mile.[8] Finally, an earth radius of 6,371 km (K = 111.195 km/degree) corresponds to the mean radius of the Clarke spheroid of 1866.

Until such time as the FCC so specifies, the author suggests using the mean Clarke spheroid value of 6,371 km (K = 111.195 km/degree).[9]

AZIMUTH CALCULATIONS

Because the FCC Rules are silent on how azimuth, or bearing, calculations are to be performed, both the flat-earth and spherical-earth methods are commonly used. The flat-earth method determines azimuth using the arctangent of the right triangle defined in the FCC method. The spherical-earth method uses standard

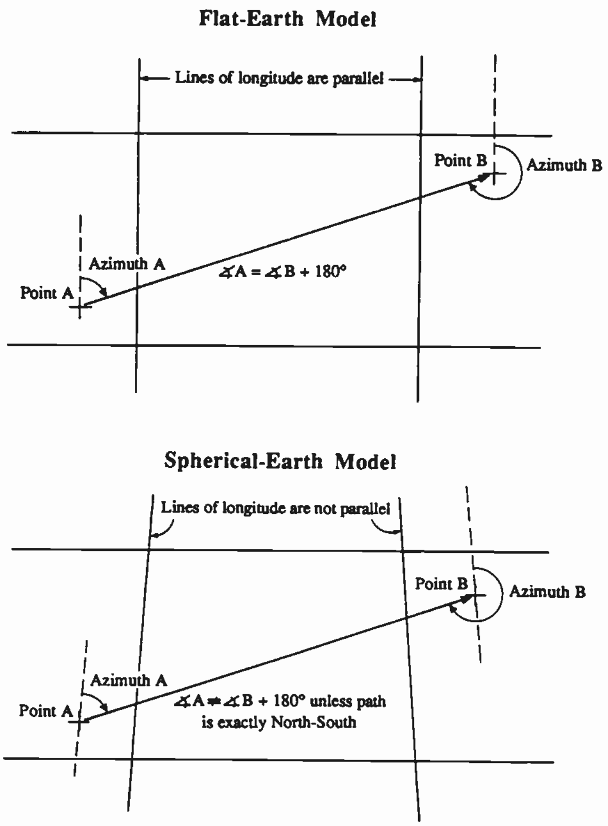


Figure 1.6-4. Azimuth calculations.

spherical trigonometry. It is the author's recommendation that azimuth always be determined using the spherical-earth method, even when the FCC flat-earth method is used to determine distance. The flat-earth method will be in error by up to two degrees at distances approaching 500 km, whereas the spherical-earth azimuth will be correct within about 0.1 degrees at such distances.[10] Figure 1.6-4 shows why this is true. Azimuths determined using the arctangent of the right triangle defined by the FCC flat-earth method assume that lines of longitude are parallel, whereas they are not. This is why the forward and back (reciprocal) azimuths using the spherical-earth method will generally not be exactly 180° apart, whereas the forward and back azimuths using the flat-earth method are always 180° apart.

The formula for determining azimuth by the spherical-earth method is:

$$C = \cos^{-1}\frac{\sin \text{Lat}_2 - \sin \text{Lat}_1 \cos(D/K)}{\sin(D/K)\cos \text{Lat}_1} \quad [1.6.8]$$

$$\text{if } \sin(\text{Long}_2 - \text{Long}_1) < 0, \text{ Azimuth} = C \quad [1.6.9]$$

$$\text{if } \sin(\text{Long}_2 - \text{Long}_1) = 0, \text{ Azimuth} = 360 - C \quad [1.6.10]$$

The ratio (D/K) is the great circle arc length *in degrees,* and is obtained from equation [1.6.7]. It should be noted that the Canadian Department of Communications Rules and the U.S.-Mexico FM Agreement *do* specify that azimuth is to be calculated using the spherical-earth method and, further, that for Canadian azimuth calculations, the bearing is to be rounded to the nearest degree.[11] It should also be noted that the radius of the sphere is irrelevant for azimuth calculations.

Clarke Spheroid Versus WGS Ellipsoid

The original FCC distance tables were based upon the Clarke spheroid (or ellipsoid) of 1866, with a major-axis radius of 6,378.2064 km and ellipticity of 1/294.98.[12] The current edition[13] of the *American Practical Navigator* now bases Table 7, "Length of a Degree of Latitude and Longitude," on the World Geodetic System (WGS) ellipsoid of 1972. The coefficients for dLat and dLong for this ellipsoid are:

$$\text{dLat} = 111.13292 - 0.55982\cos(2L) + 0.001175\cos(4L)\ldots \quad [1.6.11]$$

$$\text{dLong} = 111.41282\cos(L) - 0.0935\cos(3L) + 0.000118\cos(5L)\ldots \quad [1.6.12]$$

The difference between the Clarke 1866 and WGS 1972 trigonometric series is inconsequential when coordinate data bases are maintained only to the nearest second and distances are rounded to the nearest kilometer or one-tenth kilometer. Over a zero to 90 degree arc of latitude, at one-degree increments, the RMS difference between the two arc length formulas is only 0.0041% for latitude and 0.0033% for longitude. For this reason the FCC decided not to amend the formulas given in Section 73.208(c) of its Rules from the Clarke 1866 coefficients to WGS 1972 coefficients.

Rounding Practices

There continues to be an inconsistency in rounding practices between the FCC FM and TV Rules. For FM distance calculations, Section 73.208(c)(8) specifies rounding to the nearest kilometer. For TV distance calculations, Section 73.611(d) specifies rounding to the nearest one-tenth kilometer. Persons performing distance calculations must, therefore, be mindful of the different rounding criteria between the FM and TV Rules. The AM Rules and the Broadcast Auxiliary Rules are silent on distance rounding practices. The Broadcast Auxiliary Rules also do not prescribe how distances are to be calculated, even though portions of those Rules include minimum distance requirements (for example, Subpart G, *Low Power TV, TV Translator, and TV Booster Stations,* specifies certain minimum distances between LPTV/TV Translator stations and land mobile stations; and Subpart I, *Instructional Television Fixed Service,* specifies certain search distances for interference studies). The FCC was urged to make its distance calculation rules applicable to all Part 73 and Part 74 services in the Docket 86-144 comments, but again did not do so. Calculations regarding distances between TV Translator and LPTV stations should, presumably, use the flat-earth method specified in the FCC's FM Rules.

NAD27/NAD83 Datums

The Federal Geodetic Committee is in the process of converting all maps in the United States from the 1927 North American Datum (NAD27), which is based upon the Clarke spheroid of 1866, to the 1983 North American Datum (NAD83). NAD83 differs from NAD27 in that it is referred to the earth's center of mass, making it fully compatible with satellite systems for position determination. The FCC has indicated it will eventually convert to NAD83 to maintain accuracy and consistency with other government agencies. NAD83 is based upon the Geodetic Reference System ellipsoid of 1980 (GRS 1980), with a major axis of 6,378.135 km and an ellipticity of 1/298.26. These are the same parameters as for the WGS 1972 ellipsoid, and the trigonometric series for the one-degree arc length formulas are identical to those for WGS 1972.[14]

Current 7.5 minute topographic quadrangle maps published by the U.S. Geological Survey specify NAD27 as the reference datum, in the lower left corner of the map. At some point in the future, the U.S. Geological Survey will begin issuing maps based on the NAD83 datum. Initial attempts at printing dual-sided (NAD27/NAD83) maps proved impractical.[15]

In order to prevent intermixing of numerical information using two different map datums, the FCC has stated that the following procedures will be in effect until further notice:[16] Applicants must continue to furnish coordinates based on NAD27.

- Applicants who have filed applications with coordinates based on the NAD83 datum must provide NAD27 coordinates.
- Until further notice, the FCC will continue to specify NAD27 coordinates in its databases, authorizations, notifications, forms, and rules.
- The FCC will issue further guidance on the conversion to NAD83 as more informationbecomes available.

An exception to the above policy is FCC Form 854, *Application for Antenna Structure Registration*, which allows reporting tower coordinates in either the NAD27 or the NAD83 datums.

Conversion Between Datums

The National Geodetic Survey (NGS) has developed an algorithm known as the North American Datum Conversion program (NADCON). The program is bi-directional, meaning that it will convert NAD83 coordinates to NAD27 coordinates, as well as NAD27 coordinates to NAD83 coordinates. It also applies in Alaska, Hawaii and Puerto Rico. But, most importantly, it is the conversion algorithm recommended by statute. The *Federal Register* dated August 10, 1990, Volume 55, Number 155, Page 32681 stated:

The intent of this notice is to standardize a horizontal datum transformation method when a mathematical transformation is desired. FGCC [Federal Geodetic Control Committee] selected the method incorporated in the software identified as NADCON. It is not the intent of the notice to declare when to use a datum transformation or by what method but only to declare that when a mathematical transformation is appropriate, NADCON is recommended.

Thus, while use of NADCON is not mandatory, use of conversion algorithms other than NADCON may result in a rounding error that will cause the results to be inconsistent with those obtained by the FCC or by the Federal Aviation Administration (FAA), which do use NADCON.

NADCON uses a minimum-curvature approach to transforming between the two datums, and comes with four reference files: a large file for the contiguous United States, and smaller files for Alaska, Hawaii and Puerto Rico. By having a separate set of polynomial coefficients for each 7.5 minute topographic map, NADCON achieves an improved conversion accuracy of approximately ±0003 seconds in latitude and ±0.0005 seconds in longitude. Because of the size of these data files, programmable calculators are unable to implement the program, but today's personal computers are certainly capable of loading the NADCON data files.[17]

The NADCON software (Version 2.1) and data files are provided on two 3.5-inch disks for IBM-compatible personal computers with a math co-processor chip. The source code can be re-complied for use with other computers. A hard drive with at least 1.5 Mb of RAM is recommended.

For a more detailed discussion of the NAD27/NAD83 DATUMS, see *NAD83: What It Is and Why You Should Care*, by the author, published in the Proceedings of the 1994 SBE Engineering Conference.

USE OF GPS TO DETERMINE COORDINATES

With the widespread availability of low-cost Global Positioning System (GPS) receivers, there is a temptation to use such devices to determine latitude and longitude. Unless a considerably more expensive differential GPS receiver is used, the reported coordinates may not meet FCC accuracy requirements, which require coordinates to be reported to the nearest second of latitude and longitude. Non-differential GPS receivers will provide coordinates only accurate to ±100 meters, or approximately ±3 seconds, whereas differential GPS receivers provide coordinates accurate to ±1 meter, or approximately ±0.05 seconds.[18]

A further potential source of error exists because most GPS receivers have user-driven menus allowing display of coordinates in either the NAD27 or NAD83 datum. However, consumer grade GPS receivers are unlikely to use the NADCON conversion, and there is always the possibility that the user does not realize which datum the GPS receiver has been set to display. On more than one occasion the author has discovered parties reporting GPS-derived coordinates as NAD27 when in fact the GPS receiver had been programmed to display NAD83 coordinates. If NADCON is then used to convert the coordinates from NAD27 to NAD83, in the mistaken belief that the reported coordinates were NAD27, a bogus set of coordinates, corresponding to neither datum, are then created. Therefore, persons using GPS receivers to determine coordinates should ensure they are aware of which datum their receiver is displaying, and they should only use a GPS receiver capable of differential corrections.

REFERENCES

1. Docket 80-90 *Report and Order*, May 26, 1983.
2. *Ibid*, Page 29, Footnote 35.
3. ANSI/IEEE Standard 268-1982, *Metric Practice*, Page 31, Note 14.
4. Docket 86-144 *Second Report and Order*, September 10, 1987.
5. Personal correspondence between the author and Mr. Adam W. Mink, Chief, Hydrography and Navigation Department, Defense Mapping Agency, Washington, D.C. 20315-0030. May 24, 1985.
6. *Maps For America*, Second Edition, 1981, Page 238. Published by the U.S. Department of the Interior, Geological Survey National Center, Reston, Virginia 22092.
7. Broadcast Procedure No. 13 (BP-13), Issue 2, Broadcasting Regulation Branch, Department of Communications, Government of Canada. Effective August 6, 1987.
8. H.O. No. 9, 1981 Edition, Volume 2, Page 862, defines a nautical mile as one minute of any great circle of the earth. In 1929, the International Hydrographic Bureau proposed a standard length of 1,852 meters (exactly), which as known as the International Nautical Mile. A nautical mile of 1,852 meters implies an earth radius of 6,366.707 km [(1852 meters/minute $\times$ 60 minutes/degree $\times$ 360 degrees/circumference)/2π].
9. H.O. No. 9, Volume II, 1981 Edition, Appendix D, Page 648. The mean radius of an ellipsoid is defined as (2a + b)/3, where a is the major or equatorial radius and b is the minor or polar radius. The full precision value is 6,370.9989 km (111.1949075 km/degree), but use of the 111.195 km/degree rounded value is suggested.
10. Azimuth errors are referenced to values obtained from Andoyer-Lambert formulas. Andoyer-Lambert formulas model the earth as a true ellipsoid and are used extensively in Loran computations. The complexity of Andoyer-Lambert formulas do not warrant their routine use for FCC calculations.
11. See Canadian Department of Communications (DOC) Broadcast Procedure No. 13 (BP-13), Section 4.3, for Canadian calculations; see Appendix 2, Item 2, of the U.S.-Mexico FM Agreement of 1992, for Mexican calculations.
12. Ellipticity, or flattening, f, is defined as f = (1 − b/a), where a is the equatorial radius and b is the polar radius.
13. The *American Practical Navigator* is available from the National Oceanic & Atmospheric Administration (NOAA), 6501 Lafayette Avenue, Riverdell, MD 20737-1199, telephone 301/436-6990.
14. The major axis and ellipticity for WGS1972 and GRS1980 are identical, according to NOAA Technical Memorandum NOS NGS-16, "Determination of North American Datum 1983 Coordinates of Map Corners (Second Prediction)," by T. Vincenty, National Geodetic Survey, January 1987. Because the binomial theorem expansion of an ellipsoid model starts with only two constants, the major axis dimension and the ellipsoid models of the earth with the same major axis and ellipticity values must also be identical.
15. *Implementing North American Datum 1983 for the National Mapping Program (Ashaway Quadrangle)*, U.S. Department of the Interior, Geological Survey, Reston, Virginia 22092. Undated.
16. FCC Public Notice "FCC Interim Procedure for the Specification of Geographic Coordinates," March 14, 1988.
17. NADCON is available from NOAA, NGS, N/NGS12, 1315 East-West Highway, Station 9202, Silver Spring, Maryland 20910-3282, telephone 800/638-8972.
18. Footnote 51, November 30. 1995, WT Docket 95-5 (Tower Registration) *Report and Order*.

1.7

BROADCAST RELATED ORGANIZATIONS AND INFORMATION

COURTENAY SMITH BROWN
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC

INTRODUCTION

This chapter provides a general overview of entities whose activities affect the work of broadcast engineers. It is divided into five categories:

- Federal government
- State and local government
- Trade associations
- Professional associations
- Related broadcast-oriented organizations

Web site references are included to direct broadcast engineers to more information on the listed entity, including publications, conferences, committees, and other resources. The chapter concludes with a list of broadcast engineering and related periodicals.

Federal Government

U.S. Department of Agriculture, Forest Service
14th and Independence Avenue, SW
Washington, DC 20250
(202) 205-1760, Website: http://www.fs.fed.us

The U.S. Department of Agriculture (USDA) Forest Service oversees 190 million acres of public land under the domain of the National Forest System. Often, the use of a mountaintop transmitter site is administered by the Forest Service. Leases authorizing communications facilities are issued by the Forest Supervisor for the particular forest of interest. The Forest Supervisor has policy directives for permissible activities and fee schedules for communication uses utilizing the National Forests.

U.S. Department of Commerce, National Institute of Standards and Technology(NIST) Headquarters
Quince Orchard and Clopper Rds
Gaithersburg, MD 20899
(301) 975-2000, Website: http://www.nist.gov

The National Institute of Standards and Technology (NIST) is a non-regulatory agency within the U.S. Department of Commerce Technology Administration. NIST"s primary mission is to promote U.S. economic growth by working with industry to develop and apply technology, measurements, and standards.

U.S. Department of Commerce, Time and Frequency Division, Boulder Laboratories
325 Broadway
Boulder, CO 80303
(303) 497-3000, Website: http://www.boulder.nist.gov/timefreq/

The Time and Frequency Division, part of the Physics Laboratory of NIST, is located in Boulder, Colorado. Its functions include:

- Maintaining the primary frequency standard for the United States.
- Developing and operating standards of time and frequency.
- Coordinating U.S. time and frequency standards with other world standards.
- Providing time and frequency services for United States clientele.
- Performing basic and applied research in support of improved standards and services.

Precise time and frequency information is needed by radio and television stations, electric power companies, telephone companies, air traffic control systems, participants in space exploration, computer networks, scientists monitoring data of all kinds, and navigators of ships and planes. These users need to compare their own timing equipment to a reliable, legally traceable, internationally recognized standard. NIST provides this standard for civilian users in the United States.

Broadcast services include shortwave radio signals from NIST radio stations WWV and WWVH; 60 kHz signals from WWVB; and time signals carried by satellites. Precise frequency references are available from WWV, WWVH, WWVB, and Loran-C. Time signals are also available using telephone voice and data lines and the Internet. Time synchronization can also be accomplished via Global Positioning Satellites (GPS).

The division offers a fee-supported Frequency Measurement Service (FMS) that allows users to make accurate, NIST-certified frequency calibrations at their site, rather than sending their oscillators to NIST or elsewhere for calibration. Broadcast engineers requiring more information on this service can contact:

John A. Wessels
Time and Frequency Division
Mailcode 847.00
325 Broadway, Boulder, CO 80303

voice: (303) 497-5634, fax: (303) 497-6461
email: wessels@boulder.nist.gov

National Telecommunications and Information Administration
14th Street and Constitution Avenue N.W.
Washington, D.C. 20230
Phone: (202) 482-1835, Website: http://www.ntia.doc.gov

The National Telecommunications and Information Administration (NTIA) of the Department of Commerce is the Executive Branch agency principally responsible for domestic and international telecommunications and information technology issues. With roughly $65 million available, NTIA works to spur innovation, encourage competition, help create more jobs, and provide consumers with more choices and better quality telecommunications products and services at lower prices.

NTIA's mission is to promote the development of an advanced telecommunications and information infrastructure that efficiently serves the needs of all Americans, creates job opportunities for American workers, and enhances the competitiveness of U.S. industry in the global marketplace.

NTIA addresses a broad range of telecommunications issues and concerns. Many efforts cross-cut across the agency, drawing on telecommunications policy expertise, radio wave propagation knowledge, spectrum engineering, and lessons learned from actual applications.

U.S. Department of the Interior, Bureau of Land Management
1849 C Street, N.W.
Washington, D.C. 20240
(202) 452-5125, Website: http://www.blm.gov

The Bureau of Land Management oversees the use of 360 million acres of "public domain" lands that are not under the domain of the USDA Forest Service. Broadcaster use of these lands for transmitter sites is coordinated by the State Director of the Bureau of Land Management.

U.S. Department of the Interior, U.S. Geological Survey National Center
12201 Sunrise Valley Drive
Reston, VA 20192
Phone: (703) 648-4000, Website: http://www.usgs.gov

The U.S. Geological Survey provides the nation with information to describe and understand the Earth. This information is used to:

- Minimize loss of life and property from natural disasters
- Manage water, biological, energy, and mineral resources
- Enhance and protect the quality of life
- Contribute to wise economic and physical development

In support of the U.S. Geological Survey's mission to provide information about the Earth and its physical resources, the National Mapping Program (NMP) provides geographic, cartographic, and remote sensing information, maps, and technical assistance, and conducts related research responsive to national needs.

The NMP also makes available maps, images, spatial data, remote sensing data, and related information; provides assistance in selecting, acquiring, and using geographic and cartographic products; and designs, prints, and distributes maps of the National Atlas. It coordinates Federal topographic mapping and digital cartographic activities and provides leadership in the development and advancement of surveying and mapping technology. The NMP is administered through the USGS National Mapping Division (NMD).

Topographic maps usually used in the prediction of coverage and other engineering studies, which require accurate information about the position and elevation of terrain features may be obtained from the U.S. Geological Survey. Refer to your local phone book's yellow pages for the nearest commercial dealers that sell USGS maps and contact them directly for pricing and ordering information. For more information or ordering assistance, call 1-800-HELP-MAP, or write:

USGS Information Services
Box 25286
Denver, CO 80225

Department of Labor, Occupational Safety and Health Administration
200 Constitution Ave., N.W.
Washington, DC 20210
(202) 219-8151 (OSHA), Website: http://www.osha.gov

Broadcast engineers should be aware that two federal laws—the Federal Labor Standards Act and Occupational Safety and Health Act—empower the Department of Labor (DOL) to regulate work place safety standards and the wages and hours of employment for broadcast employees. Occupational safety standards are enforced by the Occupational Safety and Health Administration (OSHA) of the Department of Labor (DOL), while the minimum wage and hours of employment are enforced by DOL's Wage and Hour Division.

Applicability of OSHA regulations depend upon hazards, while Wage and Hour Division regulations encompass many factors, such as the station's geographic location, the number of station employees, and the type of work they perform. Broadcast engineers should consult their station's attorney to determine what regulations apply to them and how to abide by them.

Department of Transportation, Federal Aviation Administration
400 7th Street, S.W.
Washington, D.C. 20590
(202) 426-4000, Website: http://www.faa.gov

When construction or alteration of a broadcast tower is proposed, the Federal Aviation Administration (FAA) conducts an aeronautical study to determine the potential impact that the proposal may have on the navigable airspace.

Broadcasters are required to notify the FAA of new proposals or alterations to existing towers. The obstructions standards in Title 14 of the Code of Federal Aviation Regulations (Part 77) are applied to determine the effect the proposal would have on aeronautical operations. It may be necessary for broadcasters to amend their proposal by modifying the tower height or location in order to eliminate a *determination of hazard to air navigation.* When an aeronautical study results in a *determination of no hazard to air navigation,* the FAA will recommend the marking and/or lighting for that structure, if appropriate. The Federal Communication Commission's license will contain the marking and/or lighting recommended by the FAA.

Conspicuity is achieved only when all recommended lights are working. Therefore, it is important that any outage be corrected as soon as possible. However, the FAA must receive notice immediately of any failure or malfunction that lasts more than 30 minutes and affects a top light or flashing obstruction light regardless of its position.

Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460
(202) 260-2090, Website: http://www.epa.gov

The Environmental Protection Agency protects public health and safeguards and improves the natural environment—air, water, and land—upon which human life depends. EPA's purpose is to ensure that:

- Federal environmental laws are implemented and enforced fairly and effectively.
- Environmental protection is an integral consideration in U.S. policies concerning economic growth, energy, transportation, agriculture, industry, international trade, and natural resources.
- National efforts to reduce environmental risk are based on the best available scientific information.
- All parts of society—business, state and local governments, communities, and citizens—have full access to information so that they can become full participants in preventing pollution and protecting human health and the environment.

When broadcasters construct station facilities such as antenna towers or large satellite earth stations they should be aware that these activities may fall under the scope of the National Environmental Protection Act (NEPA). NEPA grants the Environmental Protection Agency (EPA) authority to regulate activities that may affect the "quality of the human environment." The FCC cooperates with EPA in enforcing provisions of NEPA that relate to telecommunication licensees.

Federal Communications Commission
1919 M Street N.W.
Washington, D.C. 20554
(202) 418-0200, Website: http//www.fcc.gov

For detailed information on the Federal Communications Commission, see Chapters 1.3, "FCC Organization and Administrative Procedures," and 1.4, "FCC Compliance and Information Bureau." Every station should have a current copy of the FCC's broadcast rules. There are two ways to obtain them:

- Order a set of the Code of Federal Regulations (CFR)—*47 CFR Parts 0–19* for the tower regulations in Part 17 and the Emergency Alert System (EAS) regulations in Part 11, and *47 CFR Parts 70–79* for the broadcast rules Part 73 and broadcast auxiliary rules Part 74—issued annually by the U.S. Government Printing Office (GPO). For payment by credit card, call (202) 512-1800, Monday–Friday, 8 a.m.–4 p.m. EST. Or fax your order to (202) 512-2233, 24 hours a day. For payment by check, write to the Superintendent of Documents, Attn: New Orders, P.O. Box 371954, Pittsburgh, PA 15250-7954. For GPO Customer Service call (202) 512-1803.
- For a continuously updated subscription (in loose-leaf form or on diskette), call the Rules Service Company at (301) 4249402 or fax your subscription request at (301) 762-7853, Monday–Friday, 8 a.m.–3:30 p.m. EST. Send written requests to: 7615 Standish Place, Rockville, MD 20855. To view the list of available Parts/Titles, visit their Web site at http://www.ruleserv.com.

National Labor Relations Board
1099 14th Street, N.W.
Washington, D.C. 20570
(202) 208-3000, Website: http://www.nlrb.gov

Broadcast engineers should be aware that the National Labor Relations Act (NRLA) protects the right of station employees to bargain collectively with management over the "terms and condition of employment." The National Labor Relations Board (NLRB), an independent federal agency, was established to enforce the right of workers to organize and engage in "concerted activity." As with the Department of Labor regulations, the extent of these rights depends upon many factors such as the type and the number of employees involved.

State and Local Governments

A number of aspects of a broadcast engineer's job are affected or controlled by state or local government agencies. It is important for engineers to have some familiarity with the laws, codes, and zoning ordinances governing such matters as building construction, electrical wiring, and fire safety. Regulations may vary from one community to another, even within the same state or county.

While there are model national codes, these codes may or may not be adopted by a state or local government. If adopted, there may be some changes from the national model. The only way to determine this is to check with the local agency or agencies having jurisdiction over the matter in question. For more information on state or local regulations, see the following:

- The broadcast station's local lawyer
- The county or city business licensing office
- The county or city building inspector or fire marshal
- A licensed local contractor who performs the type of work in question

Trade Associations

Association of Federal Communications Consulting Engineers (AFCCE)
Website: http://www.afcce.org

AFCCE was founded in 1948 as a professional association of communications engineers practicing before the Federal Communications Commission. Engineering for broadcast stations in the AM, FM, and TV services, for microwave, cellular radio, PCS, paging systems, cable systems, and for satellite facilities are some of the areas in which AFCCE members offer their professional services. Associate membership is offered to technical personnel and to other professionals sharing an interest in the technical aspects of communications.

The purpose of the Association is to aid and promote the proper federal administration and regulation of those engineering and technical phases of communications which are regulated by the Federal Communications Commission.

Association for Maximum Service Television (MSTV)
1776 Massachusetts Avenue, N.W.
Suite 310
Washington, DC 20036
Phone: (202) 861-0344, Fax: (202) 861-0342, Website: http://www.mstv.org

The Association for Maximum Service Television, Inc. (MSTV) is a national association of local television stations dedicated to preserving and improving the technical quality of free, universal, community-based television service to the American public.

In the spring of 1996, MSTV and the Consumer Electronics Manufacturers Association (CEMA) created the Model HDTV Station Project to provide hands-on education and experience for broadcasters and equipment manufacturers in the implementation of DTV. The project's station, licensed by the FCC as WHD-TV in Washington, DC, is supported by more than 260 funding television stations and by major equipment manufacturers.

Cable Television Laboratories Inc. (CableLabs)
400 Centennial Parkway
Louisville, Colorado 80027-1266
Phone: (303) 661-9100, Fax: (303) 661-9199, Website: http://www.cablelabs.com

CableLabs is a research and development consortium of cable television system operators representing more than 85 percent of the cable subscribers in the United States.

Electronic Industries Alliance (EIA)
Consumer Sector, Consumer Electronics Manufacturers Association (CEMA)
Telecommunications Sector, Telecommunications Industry
Association (TIA)
2500 Wilson Boulevard
Arlington, VA 22201-3834
Phone: (703) 907-7500, Fax: (703) 907-7501, Website: http://www.eia.org

For more than 75 years, the Electronic Industries Alliance has been the national trade organization representing U.S. electronics manufacturers. Committed to the competitiveness of the American producer, EIA represents the entire spectrum of companies involved in the design and manufacture of electronic components, parts, systems, and equipment for communications, industrial, government, and consumer uses.

Engineering activity includes the publication of voluntary standards. Those of interest to broadcasters are standards covering transmitters, towers, microwave transmission systems, and tape cartridges. The standards work is primarily done in the National Radio Systems Committee (NRSC) and standing EIA Engineering Department Committees. The NRSC is a joint engineering committees with the National Association of Broadcasters (NAB). The EIA, Joint Electron Device Engineering Council (JEDEC, a cooperative effort of EIA and the National Electrical Manufacturers Association), and TIA Standards and Engineering Publications catalog of standards and engineering publications may be obtained from Global Engineering Documents (1-800-854-7179).

A selection of EIA voluntary standards of interest to broadcast engineers is given next, along with the date of acceptance or reaffirmation. Interested readers should contact Global for a current listing and prices, or visit EIA's Website.

- EIA-189-A, Encoded Color Bar Signal, 1976.
- EIA-200-A, Circular Waveguides, 1975.
- EIA-211-D, Processed Analog Disc Records and Reproducing Equipment, 1981.
- EIA-215, Basic Requirements for Broadcast Microphone Cables, 1958.
- EIA-219, Audio Facilities for Radio Broadcasting, 1959.
- EIA-221-A, Polarity or Phase of Microphones for Broadcasting, Recording, and Sound Reinforcement, 1979.
- EIA-222-D, Structural Standards for Steel Antenna Towers and Antenna Supporting Structures, 1987.
- EIA-225, Rigid Coaxial Transmission Lines, 1975.
- EIA-232-D, Interface Between Data Terminal

Equipment and Data Circuit Terminating Equipment Employing Serial Binary Data Interchange, 1986.
- EIA-238-B, Standards for Stylus Tips Used for Disc Phonograph Record Reproducing, 1981.
- EIA-250-C, Electrical Performance Standards for Television Relay Facilities, 1977.
- EIA-258, Semi-flexible Air Dielectric Coaxial Cables and Connectors, 1962.
- EIA-261-B, Rectangular Waveguides, 1979.
- EIA-264, Magnetic Recording Tape Cartridge Dimensions, 1962.
- EIA-288, Audio Magnetic Playback Characteristic at 7 1/2 IPS, 1963.
- EIA-295, Disc Recording Characteristic, 1982.
- EIA-297-A, Cable Connectors for Audio Facilities for RadioBroadcasting, 1970.
- EIA-355, Standard Dimension for Unrecorded Magnetic Sound Recording Tape, 1974.
- EIA-462, Electrical Performance Standards for Television Broadcast Demodulators, 1979.
- EIA-508, Electrical Performance Standards for Television Broadcast Transmitters, 1987.
- EIA-549, NRSC AM Preemphasis/Deemphasis and Broadcast Audio Transmission Bandwidth Specifications, 1990.
- EIA-560, Standard Method of Measurement for Compact Disc Players, 1989.
- TR-107, Electrical Performance Standards for FM Broadcast Transmitters, 1949.
- TR-117, Antennas and Combination of Antennas for FM Broadcasting Stations, 1949.

National Association of Broadcasters (NAB)
1771 N Street, N.W.
Washington, D.C. 200362891
Phone: (202) 429-5346, Fax: (202) 775-4981
Website: http://www.nab.org

The NAB represents the broadcasting industry before Congress, the courts, regulatory agencies, the White House, and before the general public. The NAB Science and Technology Department represents the industry before the FCC and other agencies on issues affecting spectrum management and technical regulations. The Department's mission is to preserve and improve the ability of radio and television stations to distribute services to consumers and businesses.

The Department conducts a series of technical conferences—NAB Broadcast Engineering Conference, NAB MultiMedia World, and NAB Communications & Connectivity—at the annual NAB spring convention. The department also provides timely, useful, and accurate technical information in *Radio & TV Tech-Check*, a weekly technical news fax and through numerous other NAB technical publications including, the *Broadcast Engineering Conference Proceedings.*

For a listing of NAB technical publications visit the technical publications web site at http://www.nab.org/scitech/files/techpub.asp. NAB publications may be ordered by calling 1-800-368-5644. The Science & Technology Department also maintains a *Broadcast Sites of Interest* web site at http://www.nab.org/scitech/files/broadsts.htm.

National Cable Television Association (NCTA)
1724 Massachusetts Avenue N.W.
Washington, D.C. 20036
(202) 775-3550, Website: http://www.ncta.com

The National Cable Television Association (NCTA) is the major trade association for the cable industry. The organization represents the cable television industry before Congress and Federal Agencies, in the courts, and before state regulatory agencies. NCTA represents cable systems serving more than 80 percent of the nation's approximately 64 million cable subscribers. It also represents over 100 cable program networks such as CNN, HBO, and MTV. In addition, NCTA represents the hardware suppliers and providers of other services to the industry.

Professional Associations

Broadcast Education Association (BEA)
1771 N Street, N.W.
Washington, D.C. 20036
(202) 429-5354, Website: http://www.beaweb.drg

The Broadcast Education Association (BEA) provides professional development for people who teach and research electronic media and multimedia. Academics, media professionals, and students participate in this organization, which has served the industry for more than 43 years. The BEA publishes the *Journal of Electronic Media, Feedback,* and the *Journal of Radio Studies.* It also administers 14 scholarships ranging from $1,250–$5,000, awarded to juniors, seniors and those pursuing graduate studies at BEA member schools.

IEEE Broadcast Technology Society
Institute of Electrical and Electronics Engineers (IEEE)
345 East 47th Street
New York, New York 10017
(212) 705-7900, Website: http://www.ieee.org

The IEEE's 311,000 members in 150 countries, includes engineers and scientists in electrical engineering, electronics, computers, and allied fields as well as over 40,000 students. The IEEE Broadcast Technology Society is one of the Institute's 36 societies addressing member interests.

The IEEE holds numerous meetings and special technical conferences, conducts lecture courses at the local level on topics of current engineering and scientific interest, assists student groups, and awards medals, prizes, and scholarships for outstanding technical achievement. Publications of the IEEE include the *Proceedings* (monthly), *IEEE Spectrum* (monthly), and *Directory* (annually) and more than 700 industry standards. The societies and councils publish journals, magazines, and conference proceedings.

Society of Broadcast Engineers (SBE)
8445 Keystone Crossing, Suite 140
Indianapolis, Indiana 46240
Phone: (317) 253-1640, Fax: (317) 253-0418, Website: http://www.sbe.org

With more than 5,300 members, the Society of Broadcast Engineers is the largest national organization for broadcast engineers. SBE is devoted to the professional development of its members and the field of broadcast engineering. SBE promotes communication between SBE members and provides national representation for those members before federal and state regulatory agencies, manufacturers, and the general public. SBE, through its 100 plus local chapters, holds monthly meetings and technical conferences. There are eight levels of membership in the society:

- Honorary Member
- Fellow
- Senior Member
- Member
- Associate
- Student
- Life Member
- Sustaining Member

SBE administers a certification program, recognizing four levels of engineering achievement:

- Certified Broadcast Technologist, CBT
- Certified Broadcast Engineer: Television or Radio, CBTE or CBRE
- Certified Senior Broadcast Engineer: Television or Radio, CSTE or CSRE
- Certified Professional Broadcast Engineer, CPBE

SBE offers certification in two non-RF categories:

- Certified Engineer/Audio, CEA
- Certified Engineer/Video, CEV

SBE also offers certification for the operator level:

- Certified Radio Operator, CRO
- Certified Television Operator, CTO

The SBE National Frequency Coordinating Committee was established in 1982 when the FCC asked the broadcast industry to identify local contacts for Part 74 frequency coordination. The SBE formed a national network of volunteers to develop local databases of frequencies and users to assist the FCC and Part 74 users. Voluntary frequency coordination is handled on a local basis by SBE coordinators.

The Ennes Educational Foundation Trust was founded in 1981. Each year the Trust awards scholarships to qualified individuals pursuing an education in broadcast engineering or to those pursuing continuing education in broadcast technology. The Trust presents workshops around the country, providing in-depth instruction to members, and underwrites educational publications of a technical broadcast nature.

Society of Cable Telecommunications Engineers (SCTE)
140 Phillips Road
Exton, PA 19341-1318
Phone: (610) 363-6888, Fax: (310) 363-5898, Website: http://www.scte.org

The Society of Cable Telecommunications Engineers Inc. (SCTE), is a non-profit professional organization formed in 1969 to promote the sharing of operational and technical knowledge in the field of cable TV and broadband communications. Through the efforts of both the national and local chapter levels of the organization, SCTE provides training opportunities, standards development, and certification at the technician and engineer levels. Additional information and exchange opportunities are provided through the Society's national conferences, Board of Directors, committees, newsletter, and associated trade journals. Based in Exton, Pennsylvania, SCTE has 15,000 members and over 70 national and international Chapters and Meeting Groups.

Society of Motion Picture and Television Engineers (SMPTE)
595 West Hartsdale Avenue
White Plains, New York 10607
(914) 761-1100, Website: http://www.smpte.org

The membership of SMPTE is comprised of professional engineers and technicians in motion pictures, television, and allied arts and sciences. The Society advances engineering technology, disseminates scientific information, and sponsors lectures, exhibitions, and conferences to advance the theory and practice of motion-picture and television engineering. As an accredited standards developer under the American National Standards Institute, SMPTE develops national standards for motion pictures, television, and sound associated with motion picture and television images. The Society also develops Recommended Practices and Engineering Guidelines. For copies of SMPTE engineering documents, contact their Standards Department.

SMPTE also makes available picture and sound test films and video tapes for use as standardized measuring tools, serves as administrator of the Secretariat of International Organization for Standardization (ISO) Technical Committee 36 on Cinematography, and of the U.S. Technical Advisory Groups for ISO/TC 36 and Insternational Electrotechnical Commission Subcommittee (IEC/SC) 100B on Recording.

The Society sponsors technical courses at universities on such subjects as digital television, sound techniques, laboratory processing, special effects, and lighting for technicians and students. The Society presents nine annual awards for outstanding contributions to motion-picture and television engineering.

SMPTE engineering committees include:

- Audio Recording and Reproduction Technology
- Film Technology

- Laboratory Services Technology
- Projection Technology
- Video Recording and Reproduction Technology
- Television Production Technology
- Television Signal Technology
- Multimedia Technology and Packetized Television Technology

Publications of the Society include the monthly *Journal,* books of papers presented at the Society's Conferences, and other books on motion-picture and television technology.

RELATED ORGANIZATIONS

Advanced Television Systems Committee (ATSC)
1750 K Street, N.W.
Suite 800
Washington, DC, 20006
Phone: (202) 828-3130, Fax: (202) 828-3131, Website: http:\\www.atsc.org

The ATSC, comprised of over one hundred member corporations, associations, and research and educational institutions around the world, was established in 1983. ATSC is developing voluntary standards for the entire range of advanced television systems, and is also developing digital television implementation strategies and creating a certification program for television sets, computers, and other consumer video devices.

On December 24, 1996, the U.S. Federal Communications Commission mandated use of the ATSC DTV Standard for digital terrestrial television broadcasting in the United States, except for the specific video formats spelled out in the Standard, which are being implemented as a voluntary industry standard.

American National Standards Institute (ANSI)
11 West 42nd Street
New York, New York 10036
Phone: (212) 642-4900, Fax: (212) 398-0023, Website: http://www.ansi.org

The Institute represents the interests of its nearly 1,400 company, organization, government agency, institutional and international members through its headquarters in New York City, and its satellite office in Washington, D.C. ANSI serves as the clearinghouse for nationally-coordinated voluntary safety, engineering, and industrial standards. The institute gives status as American National Standards to standards developed by agreement from all groups concerned, in such areas as:

- Definitions, terminology, symbols, and abbreviations
- Materials and performance characteristics
- Procedures and methods of rating
- Methods of testing and analysis
- Size, weight, volume, and rating
- Practices, safety, health, and building construction

The organization also provides information on foreign standards, and represents United States" interests in international standardization work.

American Radio Relay League (ARRL)
225 Main Street
Newington, CT 06111
Phone: (860) 594-0200, Fax: (860) 594-0259, Website: http://www.arrl.org

The American Radio Relay League's (ARRL) membership is comprised of licensed amateur radio operators in the U.S. and Canada, and others interested in amateur radio, communication, and experimentation. The league operates the National Traffic System—a nationwide message-handling network Its members serve as official relay stations, observers, phone stations, emergency coordinators, experimental stations, and bulletin stations. They operate an experimental equipment laboratory and maintain the Museum of Amateur Radio. They also sponsor contests and present awards for operating proficiency. ARRL serves as secretariat for International Amateur Radio Union. Publications include *QST* (monthly) and *Radio Amateur's Handbook* (annually).

Audio Engineering Society (AES)
60 East 42nd Street
Room 2520
New York, NY 10165
(212) 661-8528, Website: http://www.aes.org

The Audio Engineering Society (AES) is devoted exclusively to audio technology. Its membership includes engineers, scientists and other authorities throughout the world. The AES serves its members, the industry, and the public by stimulating and facilitating advances in the constantly changing field of audio. It encourages and disseminates new developments through annual technical meetings and exhibitions of professional equipment, and through the *Journal of the Audio Engineering Society,* the professional archival publication in the audio industry.

Illuminating Engineering Society of North America (IESNA)
120 Wall Street, 17th Floor
New York, NY 10005
(212) 248-5000, Website: http://www.iesna.org

IESNA is the recognized technical authority for the illumination field. For over 90 years its objective has been to communicate information on all aspects of good lighting practice to its members, to the lighting community, and to consumers through a variety of programs, publications, and services. The strength of IESNA is its diversified membership: engineers, architects, designers, educators, students, contractors, distributors, utility personnel, manufacturers, and scientists, all contributing to and benefiting from the Society.

IESNA is a forum for exchange of ideas and information and a vehicle for its members" professional development and recognition. Through its technical committees, with hundreds of qualified members from the lighting community, the IESNA correlates vast amounts of research, investigations, and discussions to guide lighting experts and laymen on research- and consensus-based lighting recommendations. Complete lists of current and available recommendations may be obtained by writing to the IESNA Publication Office.

In addition to the *IESNA Lighting Handbook,* the IESNA publishes *Lighting Design + Application* (LD+A) and the *Journal of the Illuminating Engineering Society* (JIES) as the official magazines of the Society.

The National Radio Systems Committee (NRSC)
1771 N Street, N.W.
Washington, DC 20036
voice: (202) 429-5346, fax: (202) 775-4981, Website: http://www.nab.org

The NRSC is jointly sponsored by the NAB and the Consumer Electronics Manufacturers Association (CEMA). Its purpose is to study and make recommendations for technical standards that relate to radio broadcasting and the reception of radio broadcast signals. The NRSC is a vehicle by which broadcasters and receiver manufacturers can work together towards solutions to common problems in radio broadcast systems.

Anyone who has a business interest in the technology being investigated by the NRSC is welcome to join the Committee and participate in its activities. Members of the NRSC are generally engineers, scientists, or technicians with in-depth knowledge of the subject being studied. In order to promote the free exchange of ideas during Committee work, members of the press are not allowed to attend NRSC meetings. However, members of the press are free to contact Committee chairpersons, NAB, or CEMA with general questions about meetings. NRSC meetings are held on an as-needed basis, and NRSC members must participate at their own expense.

Underwriters Laboratories (UL)
333 Pfingsten Road
Northbrook, Illinois 60062
(847) 272-8800, Website: http://web138.bbnplanet.com/

UL is the leading third-party certification organization in the United States and the largest in North America. As a not-for-profit product safety testing and certification organization, UL has been evaluating products in the interest of public safety since 1894. Additional UL testing laboratories are located in Melville, New York; Santa Clara, California; and Tampa, Florida. UL seeks "by scientific investigation, study, experiments, and tests to determine the relation of various materials, devices, products, equipment, constructions, methods, and systems to hazards appurtenant thereto or to the use thereof affecting life and property, and to ascertain, define, and publish standards, classifications, and specifications for materials, devices, products, equipment, construction, methods, and systems affecting such hazards, and other information tending to reduce loss of life and property from such hazards."

Broadcast Engineering and Related Periodicals

ABU Technical Review
Asian Pacific Broadcast Union
P.O. Box 1164, Pejabat Pos Jalan
Pantai Baru
59700 Kuala Lumpur
Malaysia
Tel: (60-3) 282-3108, Fax: (60-3) 282-5292

Audio
1633 Broadway, 45th Floor
New York, NY 10019
Tel: (212) 767-6000, Fax: (212) 767-5633

Broadcast Engineering
9800 Metcalf
Overland Park, Kansas 66212-2215
Tel: (913) 967-1737, Fax: (913) 967-1907, E-mail: be@tertec.com

BE Radio
9800 Metcalf
Overland Park, Kansas 66212-2215
Tel: (913) 341-1300, Fax: (913) 967-1905

Broadcasting & Cable
1705 DeSales Street, N.W.
Washington, DC 200364400
Tel: (202) 659-2340, Fax: (202) 429-0651

EBU Technical Review
European Broadcast Union
Case Postale 67
CH-1218 Grand-Saconnex
Geneva, Switzerland
Tel: (+41 22) 717 2111, Fax: (+41 22) 798 5897

Chilton's ECN Electronic Component News
One Chilton Way
Radnor, PA 19089
Tel: (610) 964-4343, Fax: (610) 964-4343, Homepage: http://www.chiltonco.com

Electronic Buyers' News
CMP Publications, Inc.
600 Community Drive
Manhasset, NY 110303847
Tel: (516) 562-5899, Fax: (516) 562-5325

Electronic Engineering Times
CMP Publications, Inc.
600 Community Drive
Manhasset, NY 110303885
Tel: (516) 562-5000, Fax: (516) 562-5325

Electronic Media
Crain Communications Inc.
740 North Rush Street
Chicago, IL 60611
Tel: (312) 649-5465, Fax: (312) 649-5293,
E-mail: emediach@aol.com

Electronic Design
611 Route 46 W.
Hasbrouk Heights, NJ 07604
Tel: (201) 393-6060, Fax: (201) 393-0204,
E-mail: rallen@class.org

International Broadcasting Engineer
Queensway House
2 Queensway
Redhill
Surrey RH1 1QS England

Journal of the Audio Engineering Society
60 East 42nd Street
New York, New York 10165-2520
Tel: (212) 661-2355, Fax: (212) 682-0477

Mix
6400 Hollis Street #12
Emeryville, CA 94608-1028
Tel: (510) 653-3307, Fax: (510) 653-5142

Popular Communications
76 North Broadway
Hicksville, NY 11801
Tel: (516) 681-2922, Fax: (516) 681-2926

Pro Sound News
2 Park Ave
New York, NY 10016-9301
Tel: (212) 213-3444, Fax: (212) 213-3484

QST
225 Main Street
Newington, CT 06111
Tel: (860) 665-7531, Fax: (860) 665-7531, E-mail:
2155052@mci.com

Radio Shopper & Radio Guide
511 18th Street, SE
Rochester, MN 55904
Tel: (507) 280-9668, Fax: (507) 280-9143, E-mail:
radio@broadcast.net

Radio World
5827 Columbia Pike, 3rd Floor
Falls Church, Virginia 22041-2030
Tel: (703) 998-7600, Fax: (703) 998-2966

SMPTE Journal
595 West Hartsdale Avenue
White Plains, New York 10607
Tel: (914) 761-1100, Fax: (914) 761-3115

Television Broadcast
2 Park Ave
New York, NY 10016-5601
Tel: (212) 779-1919, Fax: (212) 213-3484,
E-mail:tvbcast@psn.com

Television Technology
P.O. Box 1214
Falls Church, VA 22041
Tel: (703) 998-7600, Fax: (703) 998-2966

Video Systems
9800 Metcalf Ave
Overland Park, Kansas 66212-2215
Tel: (913) 967-1767, Fax: (913) 967-1905, E-mail:
vs@interec.com

1.8 DIGITAL AUDIO STANDARDS AND PRACTICES

CHIP MORGAN
CMBE, EL DORADO HILLS, CA AND

RANDALL HOFFNER
ABC, INC. NEW YORK, NY

INTRODUCTION

Digital audio technology is quickly supplanting analog audio technology in United States television and radio production and broadcast facilities. Like digital video, digital audio offers many advantages in production, editing, distribution and routing. Digital audio is remarkably robust and far less susceptible to degradation from hum, noise, level anomalies and stereo phase errors than analog audio. Each analog audio recording generation and processing step adds its own measure of noise to the signal, but in the digital domain audio is not subject to such noise build-up. Once audio is converted to the digital domain, there are a number of storage options beyond magnetic tape. Digital audio may be stored on magnetic, optical, or magnetooptical discs and in solid state memories. When audio samples have been reduced to a series of numbers, processing and manipulation become largely mathematical operations, and a whole new world of possibilities opens. Non-linear editing is an example of a process that cannot be done in the analog domain. Ultimately, the breakthrough technology of digital compression creates new economies in the storage and transport of digital audio and permits the broadcast of digital audio within a reasonable segment of spectral bandwidth.

The distribution and routing of audio in the digital domain present the broadcaster with new options as well, such as the capability to embed digital audio within a serial digital video signal, facilitating the carriage of video and multiple audio channels on a single coaxial cable. Although digital audio presents its own unique set of challenges, its advantages far outweigh its disadvantages.

Digital audio systems are inherently free of the hum and noise problems that can invade analog audio systems. The nature of the digital domain gives rise to a new set of considerations for the facility planner and designer. Digital audio signals operate in the multiple megahertz frequency domain that video engineers are well acquainted with, raising such considerations as signal reflections and impedance discontinuities. System synchronization is another new concept for the audio engineer. In digital audio system engineering, just as in analog audio system engineering, cognizance of the potential pitfalls to be avoided, and the application of good engineering practices, will result in facilities that function well.

DIGITAL AUDIO STANDARDS

The fundamental signal interface for digital audio systems is AES3, the standard digital audio interface jointly developed by the Audio Engineering Society (AES) and the European Broadcast Union (EBU). AES3 is accompanied by a number of ancillary standards addressing such aspects of digital audio as sample rates, system synchronization and equipment testing, and by application notes that provide guidance in the implementation of digital audio systems.

In addition to the AES/EBU interface, several other digital audio interfaces have been and continue to be used in certain applications. These include such interfaces as the Sony/Philips digital audio interface (S/PDIF), which, although it is the consumer version of the IEC 958 interface, has found its way into professional applications and the Sony developed digital audio interface SDIF-2, which is frequently used in CD mastering and multi-track recording applications.[1]

S/PDIF and SDIF are used in some confined subsystems within many digital audio plants. The requirement for multichannel interconnections between multitrack audio recorders and multitrack audio consoles resulted in the expansion of the AES/EBU interface into the Multichannel Audio Digital Interface (MADI).

AES3—AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data

The interface defined in AES3 is commonly known as the AES/EBU digital audio interface. European

[1] Since 1997, the IEC document numbering system has added 60000 to the old IEC standard number, so the official number of this standard is now IEC 60958. Because the three-digit number is more widely known, this number will be used descriptively within the chapter.

broadcasters were about a decade ahead of United States broadcasters in their initial implementation of digital audio, and the requirement for a standard professional interface led the EBU and the AES to collaborate on the development of such an interface. The result of that collaboration is AES3 (ANSI S4.40), first standardized in 1985, revised in 1992, and revised again in 1997.

The first sentence of the abstract of AES3 is quite descriptive of its scope: "The format provides for the serial digital transmission of two channels of periodically sampled and uniformly quantized audio signals on a single shielded twisted wire pair."[2] AES3 defines a digital protocol and physical and electrical interfaces for the carriage of two discrete audio channels, accompanied by various housekeeping, status and user information in a single serial digital bitstream. As its title indicates, AES3 was designed to carry linearly quantized (uncompressed) digital audio. Compressed digital audio may be carried on the IEC 958 digital audio interface. IEC 958 is identical to AES3 in protocol and electrical characteristics, but it addresses a professional implementation (AES/EBU) and a consumer implementation (S/PDIF). The AES3 interface has the capacity to carry linearly sampled digital audio at bit depths from 16 to 24, data descriptive of such factors as channel status and sample validity, along with parity checking data and user data. Total bit count per sample, including audio and housekeeping, is 32 bits. An ancillary standard, AES5, recommends use of the professional audio sample rate of 48 kHz on AES3, while recognizing the use of sample rates of 44.1 kHz, 32 kHz and 96 kHz. AES3 carries audio samples using time-division multiplexing, in which samples from each of the two represented audio channels alternate.

The data carried on the AES/EBU interface is divided into blocks, frames and subframes. An AES block is constructed of 192 frames, each frame being composed of two subframes, each subframe containing a single audio sample. A subframe begins with a preamble that provides sync information and describes what type of subframe it is, and ends with a validity bit, a user bit, a channel status bit and a parity bit.

The subframe is divided into 32 *time slots*, each time slot being one sample bit in duration. The first four time slots are filled with a 4 bit preamble. The 24 time slots following the preamble may be filled in one of two ways. As shown in Figure 1.8-1 (a), an audio sample word of up to 24 bits may fill all the time slots. Figure 1.8-1(b) illustrates that the first four time slots of the audio sample word space may be filled with auxiliary bits, which can represent user data or low-quality audio for informational or cueing purposes, for example. In all cases, the audio word is represented least significant bit (LSB) first, most significant bit (MSB) last. If digital audio words of bit depth less than the maximum are represented, the unused bits are set to logic 0. Time slots 28, 29, 30 and 31 are filled with a validity bit (V), a user bit (U), a channel status bit (C), and a parity bit (P), respectively. The subframes are assembled into frames and blocks as shown in Figure 1.8-2.

Figure 1.8-1(a). Subframe format—16–20 Bit Audio Word[3]

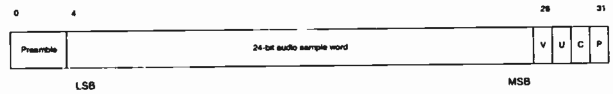


Figure 1.8-1(b). Subframe format—16–24 Bit Audio Word[4]

Each subframe begins with one of three preambles. The first subframe in the 192 frame block, a Channel 1 subframe, starts with Preamble Z. All other Channel 1 subframes in the block start with Preamble X. All Channel 2 subframes start with Preamble Y. Figure 1.8-2 represents the last frame of a block, and the first two frames of the following block. Subframe 1 of Frame 0, the first subframe of the block, begins with Preamble Z, uniquely identifying the beginning of the block. After the first subframe, the successive subframes are marked by Preamble Y and Preamble X, to identify Channel 2 and Channel 1 subframes respectively.

A frame, consisting of two 32 bit subframes, is made up of 64 bits, and the data rate of the interface signal may be readily calculated by multiplying the sampling rate times 64. In the case of the 48 kHz sample rate, the total data rate of the signal is 64 times 48,000 or 3.072 Mbps. As will be explained later, the interface employs an embedded clock signal that is twice the sample rate, making the actual frequency of this signal about 6.1 MHz.

All time slots except the preambles are encoded using biphase-mark coding to prevent the transmission of long strings of logic 0's or logic 1's on the interface, and thereby minimize the dc component on the transmission line; facilitate rapid clock recovery from the serial data stream; and to make the interface insensitive to the polarity of connections. The preambles intentionally violate the rules of biphase-mark coding by differing in at least two states from any valid biphase code to avoid the possibility of other data being mistaken for a preamble. Biphase-mark coding requires a

[2] AES3–1992 (r1997), *AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data*, New York, Audio Engineering Society, 1997, Abstract.

[3] AES3–1992 (r1997), *AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data*, Figure 1.

[4] AES3–1992 (r1997), *AES Recommended Practice for Digital Audio Engineering—Serial transmission format for two-channel linearly represented digital audio data*, Figure 1.

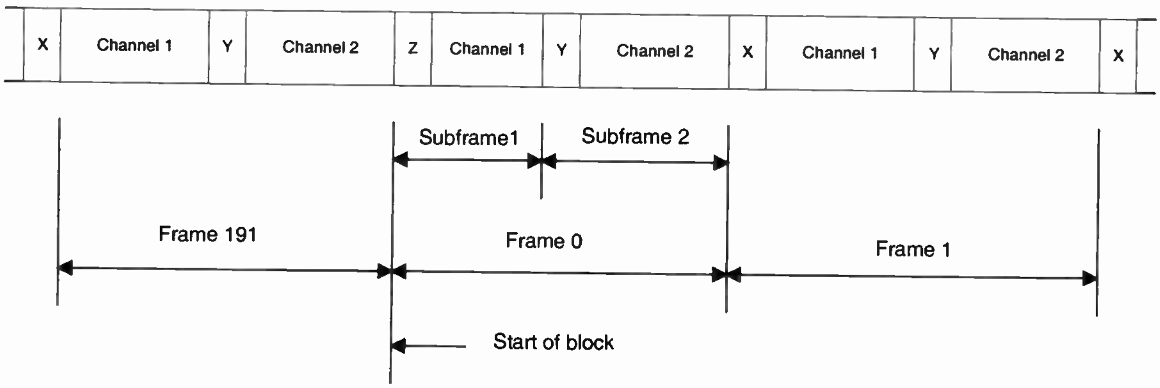


Figure 1.8-2. AES-3 Block and Frame Structure[5]

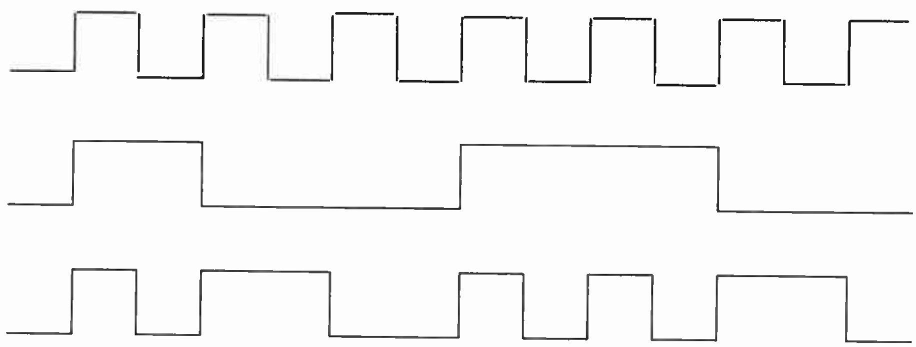


Figure 1.8-3. AES Channel Coding[6]

clock that runs at twice the sample rate of the data being transmitted, and each bit that is transmitted is represented by a symbol that is composed of two binary states. Figure 1.8-3 illustrates these relationships.

The top sequence of Figure 1.8-3 illustrates the interface clock pulses, running at a speed twice the source-coded sample rate. The middle sequence shows the source coding, which is the series of pulse code modulated (PCM) digital audio samples. The bottom sequence shows how the source coded data is represented in biphase-mark coding. In biphase-mark coding, each source coded bit is represented by a symbol that is composed of two consecutive binary states. The first binary state of a biphase-mark symbol is always different from the second state of the symbol preceding it. A logic 0 is represented in biphase-mark coding by a symbol containing two identical binary states. A logical 1 is represented in biphase-mark coding by a symbol containing two different binary states. This relationship may be seen by examining in the first full source coding bit at the left in the figure, which is a logic 1. Note that the duration of this bit is two clock pulses. Because the symbol immediately before it ended with a logic 0, the biphase-mark symbol representing it begins with a logic 1. As the bit to be transmitted is a logic 1, the second state of the biphase-mark symbol representing it is different from the first, a logic 0. The second source coded bit to be transmitted is a logic 0. Its first biphase-mark binary state is a logic 1, because the immediately previous state was a logic 0, and the second state is also a logic 1. The fact that the first binary state of a biphase-mark signal is always different from the last binary state of the previous symbol ensures that the signal on the interface does not dwell at either logic 0 or logic 1 for a period longer than two clock pulses. Because biphase-mark coding does not depend on the absolute logic state of the symbols representing the source coded data, but rather on their relative states, the absolute polarity of a biphase-mark coded signal has no effect on the information transmitted, and the interface is insensitive to the polarity of connections.

---

[5] AES3-1992 (r1997), *AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data,* Figure 1.8-2.

[6] AES3—1992 (r1997), *AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data,* Figure 1.8-3.

The last four time slots in a subframe are occupied by various housekeeping and user data. The validity bit (V) indicates whether the audio sample word is suitable for conversion to an analog audio signal. The channel status bit (C) from each subframe is assembled into a sequence spanning the duration of an entire AES3 block, and these 192 bit blocks of channel status data describe a number of aspects of the signal. Examples of channel status data include the length of audio sample words, sampling frequency, number of audio channels in use, emphasis information, consumer or professional interface implemented, audio or data being transmitted on the interface and a host of other possible information.

The 192 bit channel status bits (per block) are subdivided into 24 byte units. There is a separate channel status block for each audio channel, so channel status may be different for each of the audio channels. User data, or U-bits, may be used in any way desired. The parity bit (P) facilitates the detection of data errors in the subframe by applying even parity, ensuring that time slots 4–31 carry an even number of logic 1's and logic 0's.

The electrical interface specified by AES3 is a two-wire balanced signal. The AES interface was devised by audio engineers, with the intent to create a digital audio signal that could be carried on the same balanced, shielded, twisted pair cables and XLR-3 type connectors that are used for analog audio signals. The specified source impedance for AES3 line drivers and the specified input impedance for AES3 line receivers is 110 Ω, which is the approximate characteristic impedance of shielded twisted pair cable as used for analog audio. The permitted signal level on the interface ranges from 2–7 V peak-to-peak.

The balanced, twisted pair electrical interface can give rise to some problems in implementation. XLR-type connectors and audio patch panels, for example, are not impedance matched devices. This is of no import when the highest frequency of interest is 20 kHz, but it can cause serious problems when a 6 MHz signal must be passed. These considerations, plus the familiarity of television engineers with unbalanced coaxial transmission of 5 MHz analog video, generated the requirement for standardization of an unbalanced, coaxial electrical interface for the AES3 signal. Such an electrical interface is standardized in SMPTE 276M, which describes carriage of the AES/EBU interface on standard 75 Ω video cable using BNC connectors, at a signal level of 1 V peak-to-peak. The fact that the 110 Ω balanced and 75 Ω unbalanced signal formats coexist in many systems frequently presents the requirement to translate between these two signals. Devices to perform such translations are available for purchase, and SMPTE 276M has an informative annex explaining how to build them.

AES2–id is an information document containing guidelines for the use of the AES3 interface. *AES3–id* is an information document containing descriptive information about the unbalanced coaxial interface for AES3 audio.

AES5—AES Recommended Practice for Professional Digital Audio—Preferred Sampling Frequencies for Applications Employing Pulse-Code Modulation

This companion document to AES–3 contains the recommended digital audio sample rate for signals to be carried on the interface. The professional digital audio sample rate of 48 kHz is recommended, with recognition given to the use of the compact disc sample rate of 44.1 kHz, a low bandwidth sample rate of 32 kHz, and, new for the 1998 review, the use of a 96 kHz sampling frequency for applications requiring a higher bandwidth or more relaxed anti-alias filtering. SMPTE EG 32, engineering guideline on AES/EBU audio emphasis and sample rates for use in television systems, also recommends that the 48 kHz sample rate be used.

Variations on these sample rates are encountered. Varispeed operation requires the ability to adjust these sample rates by about ±12%, and of course accommodation to 59.94 Hz video requires operation at 48 kHz/1.001.

AES10—AES Recommended Practice for Digital Audio Engineering—Serial Multichannel Digital Audio Interface (MADI)

MADI is a multichannel digital audio interface that is based on AES3. It is designed for the carriage of up to 56 audio channels on a single coaxial cable or optical fiber. MADI preserves the AES3 subframe protocol except for the preamble. A MADI frame is composed of 56 channels, which are analogous to AES3 subframes. Each MADI channel contains 32 time slots, as does an AES3 subframe. The first four time slots contain synchronization data, channel activity status (channel on/off), and other such information. The following 28 time slots are filled in the same way as in an AES3 subframe—24 audio bits, followed by a V bit, a U bit, a P bit and a C bit.

The MADI coaxial cable interface is based on the on the fiber distributed digital interface (FDDI) standardized in ISO 9314, for which chip sets are available. Data is transmitted using non-return-to-zero inverted (NRZI), polarity-free coding and a 4–5 bit encoding format, in which each channel's 32 bits are grouped into 8 words of 4 bits each, and each 4 bit word is then encoded into a 5 bit word. The data rate on the interface is a constant 125 Mbps, with the payload data rate running between approximately 50 and 100 Mbps, depending on the sample rate in use. Sample rates may vary from 32 to 48 kHz ±12.5%. The specified coaxial cable length for the MADI signal is up to 50 m. A standard for carriage on optical fiber is under consideration.

MADI finds frequent use in multitrack audio facilities, for example, as an interface between multitrack audio recorders and consoles. It is conceivable that the MADI interface could be transmitted over very long distances, using, for example, a synchronous optical network (SONET) circuit.

AES10–id is an information document containing engineering guidelines for the implementation and use of the MADI interface.

AES11—AES Recommended Practice for Digital Audio Engineering—Synchronization of Digital Audio Equipment in Studio Operations

This document describes a systematic approach to the synchronization of AES3 digital audio signals. Synchronism between two digital audio signals is defined as that state in which the signals have identical frame frequencies, and the timing difference between them is maintained within a recommended tolerance on a sample-by-sample basis.

AES11 recommends that each piece of digital audio equipment have an input connector that is dedicated to the reference signal. Two methods of synchronization are proposed: (a) the use of a master digital audio reference signal, insuring that all input/output equipment sample clocks are locked to a single reference; and (b) the use of the sample rate clock embedded within the digital audio program signal that is input to the equipment. Method (a) is preferred for normal studio practice, as method (b) may increase the timing error between pieces of equipment in a cascaded implementation.

The digital audio reference signal is to have the format and electrical configuration of the two-channel AES3 interface, but implementation of only the basic structure of the interface format, where only the preamble is active, is acceptable as a reference signal. A digital audio reference signal may be categorized in one of two grades. A grade 1 reference signal must maintain a long-term frequency accuracy within ±1 ppm, while a grade 2 reference signal has a tolerance of less than ±10 ppm.

AES17—AES Standard Method for Digital Audio Engineering—Measurement of Digital Audio Equipment

This standard defines a number of tests and test conditions for specifying digital audio equipment. Many of these tests are substantially the same as those used for testing analog audio equipment, but the unique nature of digital audio dictates that additional tests are necessary beyond those used for analog audio equipment.

AES18—AES Recommended Practice for Digital Audio Engineering—Format for the User Data Channel of the AES Digital Audio Interface

This standard describes a method of formatting the user data channels within the AES3 digital audio interface using a packet based transmission format.

ATSC A/52—Digital Audio Compression (AC-3) Standard, Annex B, "AC-3 Data Stream in IEC 958 Interface"

Annex B of the ATSC AC-3 Digital Audio Standard for digital television broadcast describes the carriage of compressed AC-3 elementary streams on the IEC 958 digital audio interface. The advent of digital television broadcasting will necessitate the carriage of AC-3 compressed digital audio streams between pieces of DTV equipment. An example is the interface between an AC-3 encoder and the program data stream multiplexer of a DTV transmission system.

IEC 60958 Digital Audio Interface

IEC 60958 (IEC 958) is logically identical to the AES3 digital audio interface. Electrically, it provides for both the 110 Ω balanced and the 75 Ω unbalanced interfaces. Two versions are described, a consumer version, the S/PDIF interface, in which bit 0 of the channel status word is set at logic 0; and a professional version, the AES/EBU interface, in which bit 0 of the channel status word is set at logic 1. Provision is made in the location of time slots 12–27, which are normally used to carry linear 16 bit PCM audio words, to permit some recording equipment to record and play back either linear 16 bit PCM audio or encoded data streams (compressed digital audio). The consumer implementation permits only the 32 bit mode, in which channel 1 and channel 2 subframes are simultaneously employed to carry 32 bit words. The professional implementation permits either the 32 bit mode or the 16 bit mode, in which each subframe carries a 16 bit digital audio word. The consumer implementation may carry either two channels of linear PCM digital audio, or one or more compressed audio bitstreams accompanied by time stamps. The professional implementation may carry two channels of linear PCM digital audio, two sets of compressed audio bitstreams with time stamps, or one channel of linear PCM digital audio and one set of compressed audio bitstreams with time stamps.

SMPTE STANDARDS AND RECOMMENDED PRACTICES CONCERNING THE USE OF AES DIGITAL AUDIO IN TELEVISION SYSTEMS

SMPTE 272M—Formatting AES/EBU Audio and Auxiliary Data Into Digital Video Ancillary Data Space

This standard defines the embedding of AES/EBU digital audio into the standard definition serial digital interface specified in SMPTE 259M, *10-Bit 4:2:2 Component and 4f_{sc} NTSC Composite Digital Signals—Serial Digital Interface*. With such embedding, up to eight channels of digital audio in the AES3 format may be carried on the serial digital video interface signal that travels on a single coaxial cable.

SMPTE 276M—Transmission of AES/EBU Digital Audio Signals over Coaxial Cable

This SMPTE standard defines the unbalanced 75 Ω coaxial cable electrical interface for the AES3 bitstream.

SMPTE 299M—24-Bit Digital Audio Format for HDTV Bit-Serial Interface

This standard defines the embedding of AES/EBU digital audio data into the high definition serial digital video interface specified in SMPTE 292M, *Bit Serial Digital Interface for High-Definition Television Systems.* This is the high definition counterpart to SMPTE 272M.

SMPTE RP 155—Audio Levels for Digital Audio Records on Digital Television Tape Recorders

This recommended practice describes a reference level lineup signal for use in digital audio recording on digital television tape recorders, and recommends the proper setting for the lineup signal on the recorder's digital audio level meters. The reference signal is the digital representation of a 1000 Hz sine wave, the level of which is 20 dB below the system maximum (full scale digital). Meters are to be calibrated with this signal to indicate −20 dBFS (20 dB below full scale digital).

SMPTE EG 32—Emphasis of AES/EBU Audio in Television Systems and Preferred Audio Sampling Rate

This engineering guideline recommends that no emphasis be used on digital audio recordings for television applications, and that the professional digital audio sample frequency of 48 kHz be used.

IMPLEMENTATION ISSUES

Interconnection & Cabling

The key to realizing the benefits of digital audio on a system-wide scale is a thorough understanding of the principles underlying digital signal distribution, routing and switching. There are, as explained, two electrical interfaces available for AES3 signals, and both require good engineering practices for successful implementation. Digital audio's data rate dictates that uncompressed digital audio signals occupy a bandwidth similar to that of analog video. Regardless of the electrical interface, a well-engineered interconnect requires proper match of source, destination and characteristic cable impedances. Prior to the 1992 revision of AES3, any equipment manufactured to AES3–1985 violated this principle, as that standard specified a 250 Ω load impedance for receivers and a 110 Ω source impedance for transmitters. Beginning in 1992, AES3 specifies impedance matching among transmitter, receiver and cable.

Choice of Cable

The use of the unbalanced coaxial cable interface for AES3 data transmission is often preferred by video engineers. SMPTE 276M and *AES3–id* provide guidance for using the 75 Ω unbalanced AES3 interface. Any high quality video cable will be found quite acceptable for unbalanced AES3 signals. Those engineers designing facilities dealing only with audio may prefer the use of balanced, shielded, twisted pair cables with XLR-type connectors to carry AES3 signals, but should be aware of the cable length restrictions of this implementation, and of the possibility that problems will arise from impedance mismatches at connectors and patch panels. For balanced transmission of AES3 signals, special low capacitance twisted pair cable intended especially for digital audio use is recommended over the standard twisted pair cables used for analog audio, as the higher capacitance of analog audio cable tends to distort square wave signals by rolling off the higher frequency components.

Digital Audio Distribution

The use of analog video distribution and routing equipment is not recommended for AES3 signals, as such equipment may distort AES signal shapes and risetimes, adversely affecting the decoding of the signal at receiving equipment. The spurious high frequency signal energy that may be generated by such distortions of signal shape can cause crosstalk related bit errors that are difficult to dissect and analyze. Distribution of the AES3 signal using high quality digital audio distribution amplifiers maintains the proper frequency and phase relationships, as well as signal shapes and risetimes.

System Synchronization

When possible all digital audio signals should be synchronous in order to avoid objectionable digital artifacts. In a large plant, it is necessary to provide a single master reference signal to which all interconnected systems are synchronized. The master reference, fed to all pieces of equipment, allows audio data to be re-timed and synchronized within specified tolerances.

Larger facilities, in particular, will benefit from the conversion of digital audio signals from sources without external sync capability to a standard, synchronized audio sample rate. Broadcast digital audio plants typically contain consumer and other non-synchronizable equipment which requires sample rate conversion. Audio sample rate converters perform a function similar to video standards converters, in that a dynamic low pass filter continually adjusts the offending signal's phase at the output of the converter. In some cases, the output and input sample rates can be locked together via an integer relationship in a process known as *synchronous sample rate conversion*. For example, 48 kHz and 44.1 kHz are related by the integer ratio of 160 to 147.

MADI Synchronization

It is necessary for the equipment transmitting MADI data to include timing information that the receiving equipment can extract and use for synchronization. At least one sync code must be sent per frame; a sync code consists of two consecutive 5 bit words not used in the 4 bit to 5 bit encoding scheme. The total MADI interface data rate is higher than the payload data rate required, the difference between these two rates being sufficient to include sync codes within each frame. The fiber distributed digital interface (FDDI) chip set used for MADI implementation automatically handles the required synchronizing and coding operations.

Word Clock Synchronization

SDIF-2 word clock, commonly referred to as simply *word clock*, is a square wave signal at the digital audio sample rate. Word clock is commonly used as a reference signal in small, audio-only facilities. In facilities that handle both video and audio, black burst is commonly used as the reference for both video and AES audio signal synchronization.

AES3 Synchronization

AES3 is inherently synchronous, the clock signal being readily recovered from the AES3 bitstream. However, the use of a master digital audio reference insures that all digital audio equipment in a system will be frequency and phase locked and free of cascaded timing errors, and is highly recommended by AES11. The master reference signal may come from the digital audio console in a facility on the scale of a single room, or from an external reference generator in larger facilities. The master sync signal should be sent to all equipment capable of accepting external sync signals.

Digital audio phase integrity must remain intact during the conversion of multiple audio channels between the digital and analog domains. Perfect phase synchronization requires use of an SDIF-2 word clock or an AES3 signal as the common master clock. Digital audio recording and processing equipment forces any AES3 input signal into a common AES3 frame phase. When such an AES3 frame alignment is performed, a phase error will result if there are any deviations in the frame phase of A/D converters.

When digital audio signals are transferred to a piece of equipment that is not synchronized using a master sync signal, sample rate converters must be used at the inputs to the receiving equipment to prevent clicks and pops.

Signal Routing

Asynchronous routing is the simplest and most cost-efficient method of routing digital audio. It passes digital audio signals at any sample rate, a degree of flexibility that is ideal in situations where a number of different audio sample rates are encountered. However, the lack of synchronization to a master reference makes it a poor choice for on-air applications or any other situation where frame accurate switching or editing is required.

An asynchronous router may be thought of as an electronic patching system, functioning as though simple wires were used to connect inputs to outputs. In an asynchronous system, it is imperative that the destination equipment be capable of locking to the sample rate of the signal routed to it, otherwise muting usually takes place.

The disadvantage of asynchronous routers is that their output signal is almost always corrupted when a switch is made between input signals. A switch typically results in one or more AES frames being damaged, and this may cause destination equipment to momentarily lose lock, causing muting or the generation of pops and clicks.

Synchronous routing ensures precise timing and no corruption of the data stream during switches. It is considerably more complex and costly than asynchronous routing, as it requires that a transition between two inputs be made at an AES frame boundary. All inputs to a synchronous router must be locked to a common digital audio reference. A digital audio console is essentially a synchronous router with a lot of controls.

Jitter

Jitter is short-term frequency variation in the input data stream to a digital audio device. It can result from a number of causes, including such things as the coupling of excessive noise into a transmission link. Some jitter build-up is inevitable in a system, as certain components of the system inherently generate some amount of jitter. For example, noise in the phase-locked loops that control clock frequencies in the components of the system unavoidably generates some jitter. The presence of out-of-specification jitter on a digital audio signal or clock can result in bit errors that generate clicking and popping sounds. High levels of jitter may cause a receiving device to lose lock, while a relatively small amount may have no negative effect unless present in devices performing A/D or D/A conversions. Excessive jitter is seldom a problem when only two pieces of equipment are involved, but typically builds up when three or more pieces are interconnected. Jitter may be eliminated through the use of synchronizing D/As or a common synchronization signal. Jitter on the synchronization signal itself can cause degradation of digital audio in devices locked to it.

Levels and Metering

When an analog audio signal is converted to digital, the greatest analog voltage level that may be represented digitally is called full scale digital (FSD). When quantized, this voltage level causes all digital audio bits to be set to logic 1, and this level is called 0 dBFS (full scale). This is an inflexible limit, and any excursion of the analog signal above this level will be clipped off, as the digital audio word does not have the capacity to faithfully represent it. In practice, the FSD level is often set about 1 dB above the analog clip level in an effort to assure that digital clipping never occurs.

When signals are converted between the analog and digital domains, the analog reference levels of A/D and D/A converters may be set to any number of values. If the analog reference level is improperly calibrated in any of the converters in the path, A/D and D/A conversions may result in an increase or a decrease in the level of the recovered analog signal.

Consistency in the type of digital audio metering device used, good operator training and the establishment of strict house standard reference levels and alignment practices are the best defenses when it comes to accurate audio level control.

There is no United States standard for a specific digital audio level meter. Digital audiometers are often of the instantaneous response type, with no integration time, permitting them to respond with full excursion to a peak as brief as a single digital audio sample. Contrast this with the standard volume indicator (vu meter) which is an average-responding device, and the typical peak-program meter, which does not respond with full excursion to peaks with durations less than 10 ms. Typically, digital audio metering devices display a maximum value of 0 dBFS, and reference level lineup tone is set to a designated point below 0 dBFS to accommodate peaks without digital clipping.

Figure 1.8-4 shows a representative digital audiometer, the display device of which is usually an array of light emitting diodes or other such devices. This representative meter displays a range of −40 dB to 0 dBFS, with lineup tone being calibrated at −20 dBFS.

For television applications, SMPTE RP 155 recommends adjusting the level of lineup tone to read −20 dBFS on digital audiometers used on digital videotape recorders. Other industry segments have variously used lineup tone levels of −15, −18 and −20 dBFS. These varying reference levels may cause inconsistent results when digital audio recordings are interchanged. It is therefore important to establish common digital audio reference and operating levels when exchanging digital audio recordings.

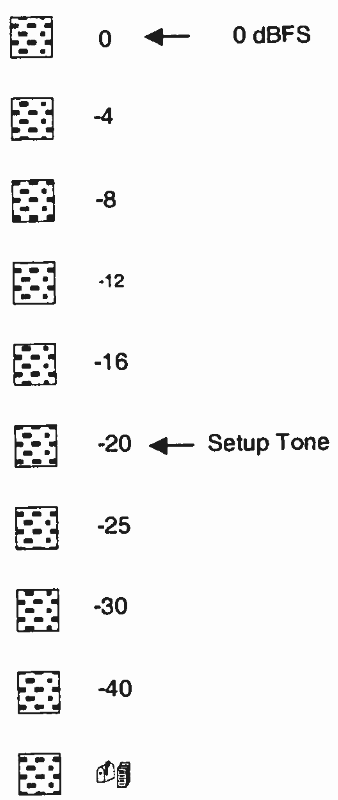


Figure 1.8-4. Representative digital audio level meter

Digital audio, with its many advantages, is not inherently susceptible to many of the problems that are encountered in analog audio systems. It does harbor some potential hazards of its own, however. With care and attention to good engineering practices in the design and maintenance of digital audio facilities, outstanding results will be realized.

REFERENCES

Standards

1. AES3–1992 (r. 1997) *AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Two-Channel Linearly Represented Digital audio Data,* New York, Audio Engineering Society, 1997.
2. AES5–1998 *AES Recommended Practice for Professional Digital Audio—Preferred Sampling Frequencies for Applications Employing Pulse-Code Modulation,* New York, Audio Engineering Society, 1998.
3. AES10–1991 (r. 1997) *AES Recommended Practice for Digital Audio Engineering—Serial Multichannel Audio Digital Interface (MADI),* New York, Audio Engineering Society, 1997.
4. AES11–1997 *AES Recommended Practice for Digital Audio Engineering—Synchronization of Digital Audio Equipment in Studio Operations,* New York, Audio Engineering Society, 1997.
5. AES17–1998 *AES Standard Method for Audio Engineering—Measurement of Digital Audio Equipment,* New York, Audio Engineering Society, 1998.
6. AES18–(1996) *AES Recommended Practice for Digital Audio Engineering—Format for the User Data Channel of the AES Digital Audio Interface,* New York, Audio Engineering Society, 1996.
7. ATSC A/52–1995 *Digital Audio Compression (AC-3) Standard* "AC-3 Data Stream in IEC 958 Interface," Annex B, Washington, Advanced Television Systems Committee, 1995.
8. IEC 60958 (1989–02) *Digital Audio Interface,* Geneva, International Electrotechnical Commission, 1989.
9. SMPTE 259M–1993 *10-Bit 4:2:2 Component and $4f_{sc}$ NTSC Composite Digital Signals—Serial Digital Interface,* White Plains, Society of Motion Picture and Television Engineers, 1993.
10. SMPTE 272M–1994 *Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space,* White Plains, Society of Motion Picture and Television Engineers, 1994.
11. SMPTE 276M–1995 *Transmission of AES/EBU Digital Audio Signals Over Coaxial Cable,* White Plains, Society of Motion Picture and Television Engineers, 1995.

12. SMPTE 292M–1996 *Bit-Serial Digital Interface for High-Definition Television Systems,* White Plains, Society of Motion Picture and Television Engineers, 1996.
13. SMPTE 299M–1997 *24-Bit Digital Audio Format for HDTV Bit-Serial Interface,* White Plains, Society of Motion Picture and Television Engineers, 1993.

Recommended Practices and Information Documents

AES2–id, 1996 *AES Information Document for Digital Audio Engineering—Guidelines for the Use of the AES Interface,* New York, Audio Engineering Society, 1996.

AES3–id, 1995 *AES Information Document for Digital Audio Engineering—Transmission of AES3 Formatted Data by Unbalanced Coaxial Cable,* New York, Audio Engineering Society, 1995.

AES10–id, 1995 *AES Information Document for Digital Audio Engineering—Engineering Guidelines for the Multichannel Audio Digital Interface (MADI) AES10,* New York, Audio Engineering Society, 1995.

SMPTE Recommended Practice RP 155–1997 *Audio Levels for Digital Audio Records on Digital Television Tape Recorders,* White Plains, Society of Motion Picture and Television Engineers, 1997.

SMPTE Engineering Guideline EG 32–1996 *Emphasis of AES/EBU Audio in Television Systems and Preferred Audio Sampling Rate,* White Plains, Society of Motion Picture and Television Engineers, 1996.

1.9
NTSC STANDARD[1]

STAFF

INTRODUCTION

Nearly every branch of science, including chemistry and psychology, contributes in some way to the reality of color television. Through chemistry, improved phosphors are continually being found for use in color picture tubes.[2] Psychology enters into the selection of lighting arrangements and picture composition to obtain desirable interpretations by the viewer. But physics plays the leading role with intense application in optics and illumination as well as in the design of electronic circuitry and components for the complete television system.

Two specialized branches of physics, namely radio and television engineering, are responsible for the electronic techniques which make color television compatible with black and white (monochrome) television.

Compatibility

When introduced in 1953, the compatible color system offered tremendous economic advantages to the home viewer and the television broadcaster. Because of compatibility, color telecasts could be seen (in monochrome) on black and white television receivers without any changes or added devices. Also, color receivers could receive monochrome as well as color telecasts. Since compatible color is transmitted over the same channels as monochrome and within the same framework of standards, the television broadcaster could utilize the monochrome system as the transmitting nucleus when installing equipment to broadcast color.

Another important advantage of the compatible color system was the part it played in the conservation of the radio frequency spectrum. Compatible color required no additional space in the spectrum. However, it employed techniques which made much more efficient use of the standards originally set up for monochrome television.

A brief review of the fundamentals of monochrome television, particularly the areas wherein specialized color methods are employed, is presented in the next few paragraphs as an aid in describing the basic color concepts.

Television: A System of Communications

Very simply, the function of the television system is to divide and subdivide the optical image into over 200,000 picture elements, each of different light intensity; convert these light elements to electrical equivalents; and transmit them in orderly sequence over a radio-frequency carrier to the television receiver.

Reversing the process at the receiver, these electrical signals are each converted to light of corresponding brightness and reassembled to produce the transmitted image on the fact of the picture tube.

Scanning

Picture elements to be transmitted in sequence are selected by a process of image scanning, which takes place in the television camera focused on the studio scene at the station. Within the camera, an electron beam in a pickup tube scans a sensitive surface containing an electrical image of the scene of action, as illustrated in Figure 1.9-1. The electron beam successively scans the image at great velocity, beginning at the upper left corner and continuing left to right in a series of parallel lines to scan the image completely. Movement of the electron beam, which can be controlled magnetically by vertical-and horizontal-deflection coils surrounding the tube, is analogous to that of the eye in reading a printed page. The speed of movement is such, however, that 30 complete image frames of approximately 500 lines each are scanned every second. Of course, at the receiver, an electron beam in the picture tube, moves with the same speed and in synchronism with the camera tube beam so that the corresponding picture elements appear in the proper relative position on the television screen (see Figure

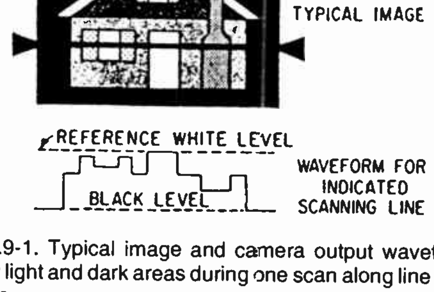


Figure 1.9-1. Typical image and camera output waveform produced by light and dark areas during one scan along line indicated by arrows.

[1] This section adapted from *Color Television, Manual for Technical Training,* RCA Corp. used with permission.

[2] In this chapter the terms CRT, kinescope and picture tube are used interchangeably.

1.9-2). Modern cameras use CCD devices in place of pickup tubes, however, the tube based explanation is the more straightforward approach.

Owing to persistence of vision and the speed of scanning, these elements appear to be seen all at once as a complete image rather than individually. Thus, the impression is one of continuous illumination of the screen and direct vision.

Scanning standards have been established in this country to assure that all television receivers are capable of receiving programs broadcast by any television station within range. The scanning pattern adhered to by manufacturers in the design of television receivers and broadcast equipment consists of 525 lines with odd-line interlaced scanning. Interlaced scanning, effective in eliminating perceptible flicker, is a method whereby the electron beam scans alternate rather than successive lines. For example, the beam begins by scanning odd-numbered lines (1, 3, 5, 7) until it reaches the bottom of the image, whereupon it returns to the top of the image to scan the even-numbered lines (2, 4, 6, 8). Thus, each scan, or field, comprises only half of the total number of scanning lines, and two fields are required to produce the 525-line frame. Each field is completed in one-half the frame time. The vertical scanning frequency is 2 × 30 or 60 Hz, and horizontal scanning frequency is 30 × 525 or 15,750 Hz. This process is illustrated in Figure 1.9-3.

Resolution and Bandwidth

The degree of resolution, or fine detail, that can be seen in a televised image depends upon the number of scanning lines used and the bandwidth of the transmitting and receiving system (see Figure 1.9-4).

The relationship between resolution and bandwidth can be seen by considering the number of picture elements that can be transmitted each second.

The standard 6 MHz broadcast channel provides a video bandwidth of approximately 4.1 MHz (the remaining bandwidth being required for a vestigial sideband plus the sound signal). Since each cycle of a sine wave is capable of conveying two picture elements (one black and one white), the maximum rate at which picture elements can be transmitted is 4,100,000 × 2, or 8,200,000 per second. Since 30 complete frames are transmitted per second, the number of picture elements per frame would be 8,200,000 ÷ 30, or 273,333 (if it were not for the retrace blanking problem, which requires interruption of the picture signal periodically by blanking pulses). Since the combination of horizontal and vertical blanking pulses requires nominally 25% of the total time, the maximum number of picture elements per frame is reduced in practice to 0.75 × 273,333, or approximately 205,000.

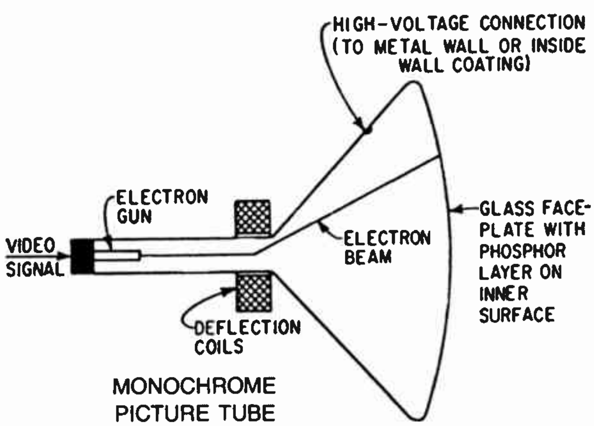


Figure 1.9-2. Diagram showing principal elements of the monochrome picture tube.

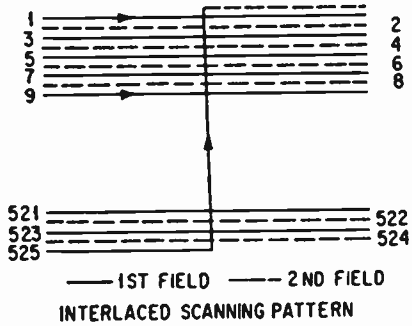


Figure 1.9-3. Diagram showing paths of the electron beam in both the pickup tube and CRT to produce the interlaced scanning pattern.

Synchronizing

In addition to the picture information, or video signals, blanking and synchronizing signals are transmitted by the television station to control the intensity and movement of the scanning beam in the CRT of the television receiver. Both these signals are in the form of rectangular pulses. Moreover, their polarity and amplitude are such that they are received as black signals and therefore do not appear on the receiver screen.

Blanking pulses eliminate the *retrace* lines which would otherwise appear between scanning lines and at the end of each field from the bottom of the picture to the top. Horizontal blanking pulses, transmitted at the end of each line, or at intervals of 1/15,750 sec,

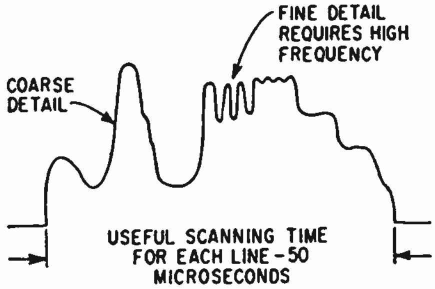


Figure 1.9-4. Diagram illustrating the relationship between picture detail and signal bandwidth.

blank the beam during retrace periods between lines. Vertical blanking pulses, transmitted at the end of each field, or at intervals of 1/60 sec, blank the beam during the time required for its return to the top of the picture. Because the vertical retrace is much slower than the horizontal, the vertical blanking periods are longer than the horizontal blanking periods. Vertical blanking pulses are about 20 lines duration, while horizontal blanking pulses have a duration of only a small fraction of a line.

Synchronizing signals keep the scanning beam of the picture tube in step with that of the camera tube. These signals consist of horizontal and vertical pulses which are transmitted within the respective blanking periods. Although the sync pulses are of the same polarity as the blanking pulses, they are of greater amplitude (blacker than black) and thus easily separated in the receiver and fed to the deflection circuits of the picture tube.

Since the vertical sync pulses are quite long compared with the horizontal sync pulses and the two are of the same amplitude, separation at the receiver is accomplished through frequency discrimination. Serrations, or slots in the vertical pulses, prevent loss of horizontal sync during the vertical blanking period.

The Monochrome Television System

The major equipment in a typical television station consists of the aural and visual units illustrated in the block diagram of Figure 1.9-5. In the visual channel, the video signal leaving the camera is passed through processing equipment which inserts the blanking and synchronizing signals and performs other functions (such as aperture compensation and gamma correction). From the processing chain, the video signal is fed to a switching system which provides for selection

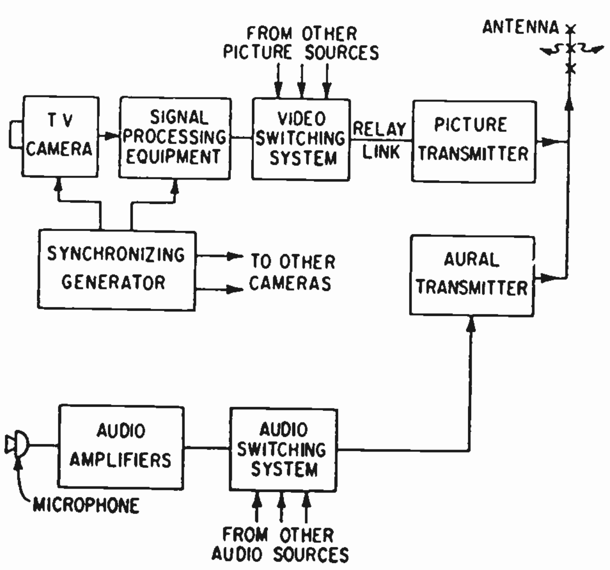


Figure 1.9-5. Simplified block diagram of the monochrome television station.

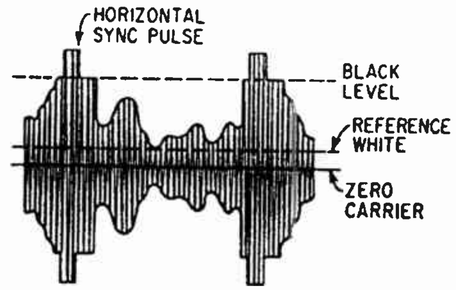


Figure 1.9-6. Waveform and radiated picture signal.

from a number of video sources. The selected signal is then sent to the visual transmitter through coaxial cable or over a microwave relay link, depending upon the distance between the television studio and transmitter. In the transmitter, the composite video signal amplitude modulates a carrier in the VHF or UHF range, which is radiated by the television antenna.

In the aural channel, the audio signal is fed from the sound sources through the switching system and to the aural transmitter. Frequency modulated output from the aural transmitter is combined with the visual output and radiated from the same antenna.

The Radiated Picture Signal

Amplitude relationships between the synchronizing pulses and the tonal gradations from white to black in the picture are represented in the waveform of the radiated picture signal. From Figure 1.9-6, it can be seen that modulation takes place in such a way that an increase in the brightness of the picture causes a decrease in carrier output power. Note that the reference white line indicated on the sketch is relatively close to zero carrier level. Also, the synchronizing pulses are in the *blacker than black* region, representing maximum carrier power. Use of a widely different range of amplitude for the sync pulses makes it possible for home receivers to separate them by a simple clipping technique.

Receiver

The basic elements of the television receiving system are illustrated in Figure 1.9-7. The radiated television signal is picked up by an antenna and fed to a tuner, which selects the desired channel for viewing. Output from the tuner is passed through an intermediate-frequency amplifier, which provides the major selectivity and voltage gain for the receiver. A second detector then recovers a video signal which is essentially the same as that fed to the visual transmitter.

The sound signal is usually taken off at the picture second detector in the form of a frequency-modulated beat between the picture and sound carriers. The sound signal is further amplified in an intermediate frequency IF stage, detected by a discriminator or ratio detector, and applied to the speaker through an audio amplifier.

Picture output from the second detector is fed to two independent channels. One of these is the video

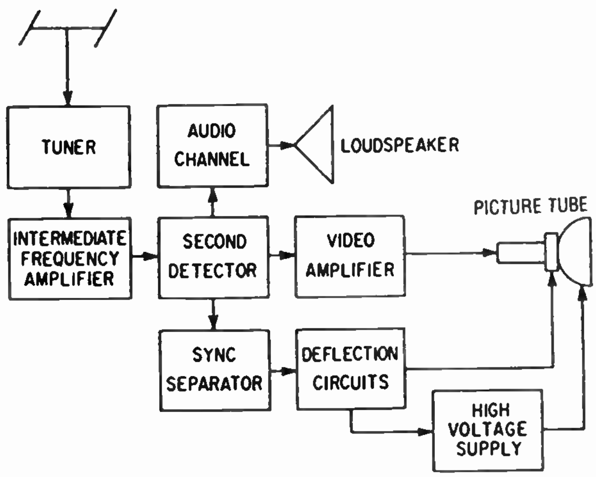


Figure 1.9-7. Block diagram of monochrome television receiver.

amplifier, which drives the electron beam in the picture tube, and the other is the sync separator, or clipper, which separates the sync pulses from the picture information. The separated pulses are then used to control the timing of the horizontal and vertical deflection circuits. The high-voltage supply, which is closely associated with the horizontal deflection circuit, provides accelerating potential for the electron beam.

The Three Variables of Color

Color is the combination of those properties of light which control the visual sensations known as *brightness, hue* and *saturation*. Brightness is that characteristic of color which enables it to be placed in a scale ranging from black to white or from dark to light. Hue, the second variable of a color, is the characteristic which enables a color to be described as red, yellow, blue or green. Saturation refers to the extent to which a color departs from white, or the neutral condition. Pale colors, or pastels, are low in saturation, while strong or vivid colors are high in saturation.

The monochrome system is limited to the transmission of images that vary with respect to brightness alone. Thus, brightness is the only attribute of a color which is transmitted over a monochrome television system. To produce a color image, therefore, provisions must be made for the transmission of additional information pertaining to all three of the variables of color. However, since the primary color process can be employed, it is not necessary to transmit information in exactly the form expressed by the three variables.

Primary Colors in Television

Experiments have proved conclusively that virtually any color can be matched by the proper combination of no more than three primary colors. While other colors could be used as primaries, red, green and blue have been selected as the most practical for color television use. Red and green combined produces yellow, red plus blue produces purple and green plus blue produces cyan or blue-green. The proper combination of all three of the primary colors produce white, or neutral, as shown at the center of the illustration. By relatively simple optical means, it is possible to separate any color image into red, green and blue, or RGB components.

Generating RGB Signals

Major components of a color television camera may have the block diagram form shown in Figure 1.9-8. Whereas the monochrome camera contains only one pickup tube, or solid-state sensor, the color camera usually contains three separate pickup devices. An

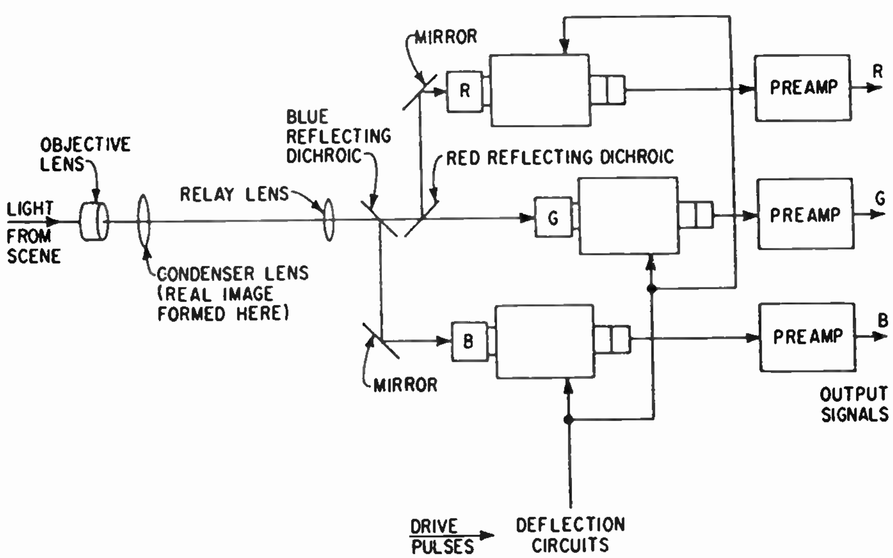


Figure 1.9-8. Simplified block diagram of the optical and electrical components of the color camera.

objective lens at the front of the camera forms a real image within a condenser lens which is located where the pickup device is usually mounted in a monochrome camera. A relay lens transfers this real image to a system of dichronic (color separating) mirrors or prisms which shunt the red and blue light to the red and blue pickup devices and permit the green to pass straight through to the green tube or sensor. In this manner, the three pickup devices produce three separate images corresponding to the RGB components of the original scene. These images are scanned in the conventional manner by common deflection circuits.

A single scanning line through the typical color image at a given point produces three separate waveforms (see Figure 1.9-9). It is important to note the correlation between these waveforms and the image at the top. The yellow shutters in the image, for example, must be produced by a mixture of red and green, and the blue signal is not required. Thus, at this interval of scanning the red and green signals are both at full value and the blue signal is at zero. The white door utilizes all three color signals. Of course, similar correlations can be seen for other parts of the image along the scanning line.

Displaying RGB Signals

RGB signals are displayed in color by the tricolor picture tube, the basic components of which are shown in Figure 1.9-10. Three electron guns produce three beams, which are independently controlled in intensity by the red, green and blue signals. These three beams are all made to scan in unison by deflection coils around the neck of the tube. The three beams converge at the screen owing to the magnetic field produced by a convergence yoke.

The phosphor screen of the color picture tube consists of an array of very small primary color dots. Approximately 1/2 in. behind the phosphor screen is an aperture mask which has one very small opening for each group of red, green and blue phosphors. Alignment of this aperture mask and screen is such that each beam is permitted to strike phosphor dots of only one color. For example, all the electrons emitted by the red gun must strike red phosphor dots on the aperture mask; they cannot strike either the green or blue dots because of the shadow effect of the mask. Likewise, the beams emanating from the other two guns strike only green or blue dots.

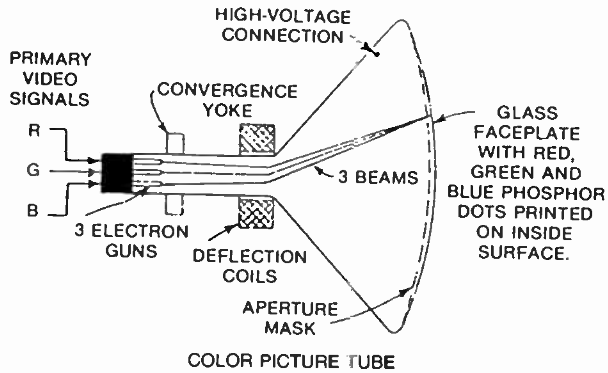


Figure 1.9-10. Diagram showing components of the three-gun picture tube.

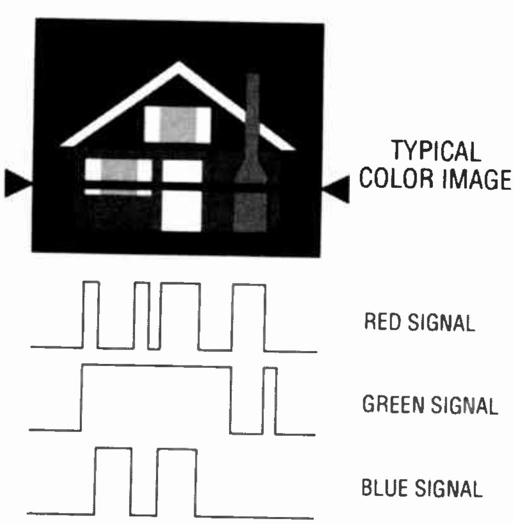


Figure 1.9-9. Typical color image and RGB waveforms.

In this way, three separate primary color images are produced on the screen of the tricolor tube. But since these images are formed by closely intermingled dots too small to be resolved at the normal viewing distance, the observer sees a full color image of the scene being televised.

ELECTRONIC ASPECTS OF COMPATIBLE COLOR TELEVISION

To achieve compatibility with monochrome television, color television signals must be processed so that they can be transmitted through the same channels used for monochrome signals, and they must also be capable of producing good monochrome pictures on monochrome receivers. Since color television involves three variables instead of the single variable (brightness) of monochrome television, an encoding process is required to permit all three to be transmitted over the one available channel. Likewise, a decoding process is required in the color receiver to recover the independent RGB signals for control of the electron guns in the color picture tube. Moreover, the process used must enable existing monochrome receivers to produce a monochrome picture from the color information (see Figure 1.9-11.)

Encoding and decoding processes used in compatible color television are based on four electronic techniques known as *matrixing, band shaping, two-phase modulation* and *frequency interlace.* It is these processes which make the color system compatible with monochrome and enable the color system to occupy the existing 6 MHz channel.

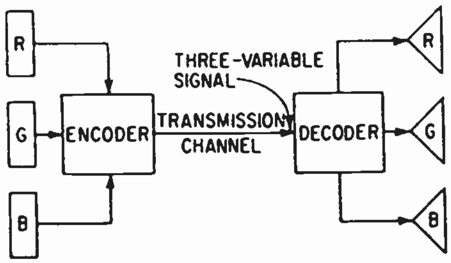


Figure 1.9-11. Encoding of the RGB signals provides a three-variable signal which can be transmitted over existing monochrome channels.

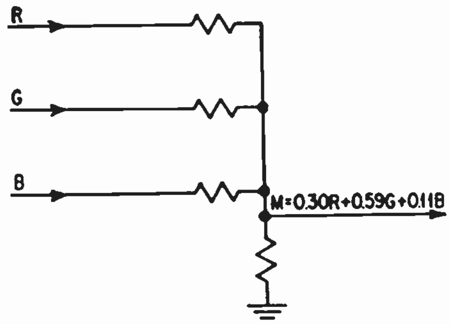


Figure 1.9-13. Diagram of resistance matrix circuit used to produce the M luminance signal.

Matrixing

Matrixing is a process for repackaging the information contained in the red, green and blue output signals from a color camera to permit more efficient use of the transmission channel. The matrix circuits which perform this function consist of simple linear cross-mixing circuits. They produce these signals, commonly designated M, I, and Q, each of which is a different linear combination of the original red, green and blue signals (see Figure 1.9-12). Specific values for these signals have been established by FCC standards.

The M-signal component, or *luminance* signal, corresponds very closely to the signal produced by a monochrome camera, and therefore is capable of rendering excellent service to monochrome receivers. The M component is obtained by combining red, green and blue signals in a simple resistor network (see Figure 1.9-13) designed to produce a signal consisting of 30% red, 59% green, and 11% blue.

The I and Q signals are *chrominance* signals which convey information as to how the colors in the scene differ from the monochrome, or neutral condition. The component I is defined as a signal consisting of 60% red −28% green and −32% blue. Minus values are easily achieved in the matrix circuits by use of phase inverters to reverse the signal polarity (see Figures 1.9-14 and 15). The Q signal is defined as 21% red −52% green and 31% blue.

It can be seen that the quantities are related so that when red, green and blue are equal, corresponding to the neutral condition, both I and Q go to zero. Thus, when the color camera is focused on an object having no color information, such as a monochrome test chart, the I-signal and Q-signal components are absent, leaving only the M component, or monochrome signal.

The matrix circuits, therefore, produce a new set of waveforms corresponding to the M, I, and Q components of the image. A comparison of the MIQ and RGB waveforms obtained from the image illustrates the correlation among the types of signals (see Figures 1.9-9 and 16). It will be seen that the M signal remains in the region between black level and reference white. The I and Q signals, on the other hand, swing positive and negative around a zero axis.

Band Shaping

The eye has substantially less acuity in detecting variations in chrominance than it has for resolving differences in brightness. This important characteristic of human vision was considered in setting up the I and Q equations because it permitted a significant reduction in the bandwidth of these signals through use of low pass filters. A bandwidth of approximately 1.5 MHz was found to be satisfactory for the I signal, which corresponds to color transitions in the range extending from orange to blue-green. For color transitions in the range from green to purple, as represented by the Q signal, the eye has even less acuity and the bandwidth was restricted to only 0.5 MHz. The M-

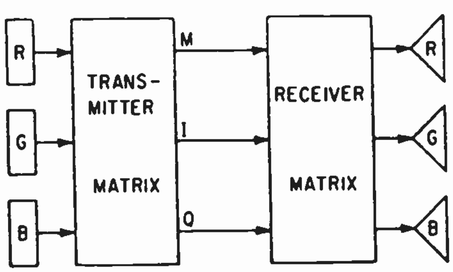


Figure 1.9-12. A part of the encoding process is the matrixing of R, G and B signals to provide M, I, Q signals.

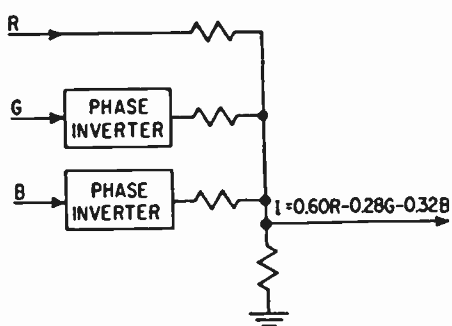


Figure 1.9-14. Diagram of 1 matrix showing phase inverters to produce minus green and blue quantities.

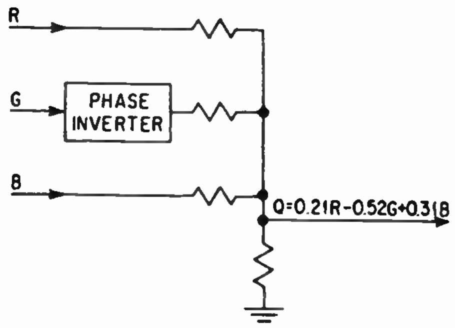


Figure 1.9-15. Diagram of the Q matrix showing phase inverter to produce required minus green signal.

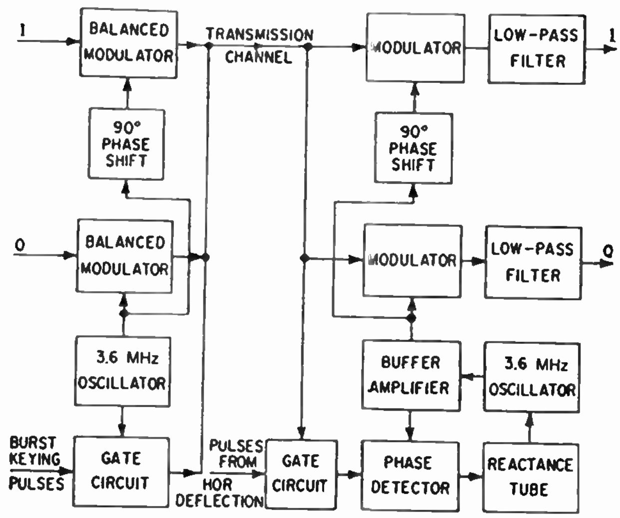


Figure 1.9-17. Simplified block diagram showing elements for transmitting and receiving the I, Q, and burst signals.

signal component, which conveys the fine details, must be transmitted with the standard 4 MHz bandwidth.

Two-Phase Modulation: Generation of Color Subcarrier

Two-phase modulation is a technique by which the I and Q signals can be combined into a two-variable signal for transmission over a single channel. This is accomplished by adding the sidebands obtained through modulation of two 3.6 MHz carriers separated in phase by 90°. The resultant waveform is the vector sum of the components. Elements of the transmitting and receiving system are shown in Figure 1.9-17. The two carriers, which are derived from the same oscillator, are suppressed by the balanced modulators. Thus, only the two amplitude modulated sidebands, 90° out-of-phase, are transmitted. At the receiving end of the system, the I and Q signals are recovered by heterodyning the two-phase wave against two locally generated carriers of the same frequency but with a 90° phase separation and applying the resultant signals through low pass filters to the matrix circuits. Typical signal waveforms are illustrated in Figure 1.9-18.

The 3.6 MHz oscillator at the receiver must be accurately synchronized in frequency and in phase with the master oscillator at the transmitter. The synchronizing information consists of 3.6 MHz bursts of at least 8 Hz duration transmitted during the back porch interval following each horizontal sync pulse as shown in Figure 1.9-19. The bursts are generated at the transmitter by a gating circuit which is turned on by burst keying pulses derived from the synchronizing generator. At the receiver, the two-phase modulated signal is applied to another gating circuit, known as a *burst separator*, which is keyed on by pulses derived from the horizontal deflection circuit. The separated bursts are compared in a phase detector with the output of the local 3.6 MHz oscillator. Any error voltage developed is applied through a smoothing filter to a varactor which corrects the phase of the local oscillator.

FCC Standard phase relationships between the I and Q signals and the color synchronizing burst are shown

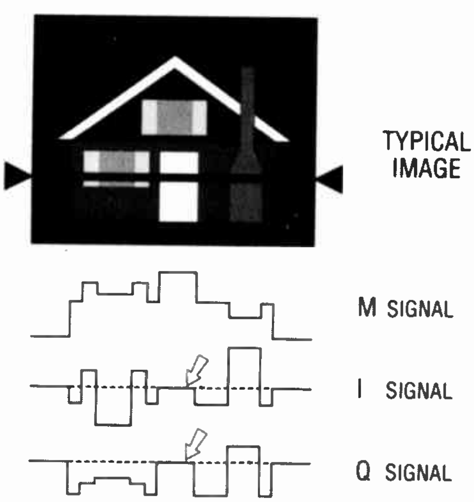


Figure 1.9-16. Typical color image and MIQ waveforms.

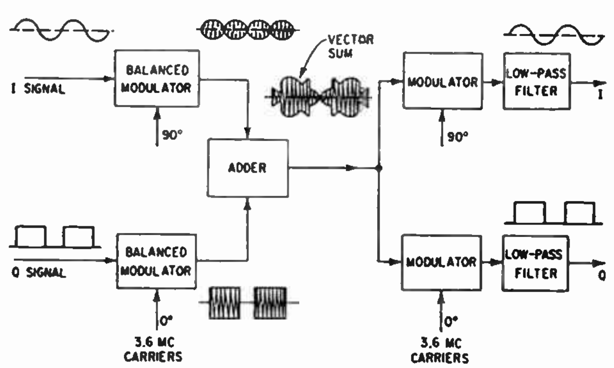


Figure 1.9-18. Representative waveforms of the separate I, Q signals and the vector sum of the suppressed carrier sidebands at the modulator output. Original I and Q signals are recovered by heterodyning in balanced modulators at receiver.

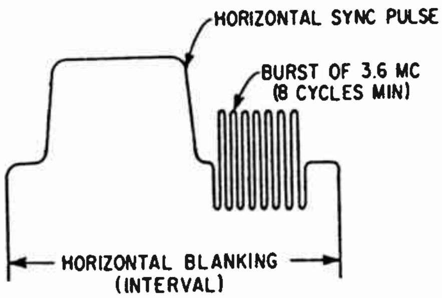


Figure 1.9-19. Diagram showing position of subcarrier burst during horizontal blanking interval.

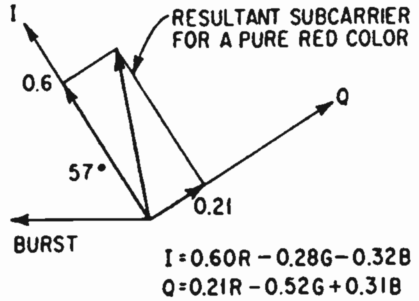


Figure 1.9-21. Vector diagram showing phase and amplitude of subcarrier for a pure red signal.

in the vector diagram of Figure 1.9-20. The I and Q signals are transmitted in phase quadrature, and the color burst is transmitted with an arbitrary 57° phase lead over the I signal.

Several interesting properties of the two-phase modulated signal are illustrated by the vector diagrams which represent the resultant signal under known transmission conditions. For example, when a pure red color of maximum amplitude is being transmitted, the green and blue components are at zero and the I and Q signals have levels of 60 and 21%, respectively. When modulated upon their respective carrier, these signals produce the resultant shown in Figure 1.9-21. The phase and amplitude shown are characteristic of pure red of maximum relative luminance. Figure 1.9-22 is a composite vector diagram showing the phase and amplitude characteristics of the three primaries and their complementary colors. This composite diagram indicates that there is a direct relationship between the *phase* of the resultant two-phase modulated signal and the *hue* of the color being transmitted. There is also a relationship (although indirect) between the *amplitude* of the resultant signal and the saturation of the color being transmitted. If the phase of the resultant subcarrier and the level of the monochrome signal both remain constant, then a reduction in the amplitude of the subcarrier indicates a decrease in color saturation. The composite vector diagram also shows an interesting symmetry between complementary colors (colors are complementary if they produce a neutral when added together); the resultants for any two complementary colors are equal in amplitude but opposite in phase.

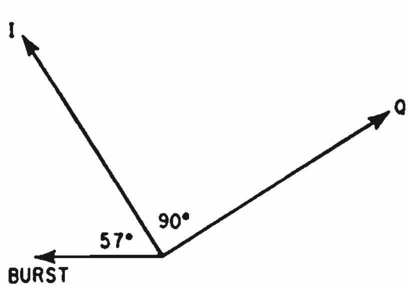


Figure 1.9-20. Diagram showing phase relationship of I, Q, and burst signals.

Frequency Interlace

Since the 3.6 MHz carriers, consisting of the I and Q sidebands, fall within the video passband as shown in the diagram of the television channel (Figure 1.9-23), they become subcarriers and can be handled in many respects like unmodulated video signals. By use of frequency interlace it is possible to add the several components of the chrominance and monochrome signals together without causing objectionable mutual interference.

The significance of the straightforward addition of signal components made possible by frequency interlace may be brought out by a study of waveforms derived from a simple color image. Figure 1.9-24 shows M, I, and Q signals after the latter two have been modulated upon 3.6 MHz subcarriers. Note that both the I-and Q-signal components are at zero during the scanning of the white door, a neutral area. Figure 1.9-25 shows the vector sum of the I and Q signals and also the complete compatible color signal formed by adding together all the components, including synchronizing pulses and color synchronizing bursts. The most significant fact about this signal is that it is still capable of providing good service to monochrome receivers, even though a modulated wave has been added to the monochrome signal component. Although the modulated wave is clearly a spurious signal with respect to the operation of the kinescope in a mono-

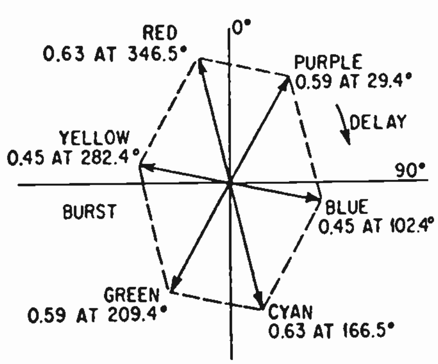


Figure 1.9-22. Composite vector diagram showing subcarrier phase and amplitude for each of six colors.

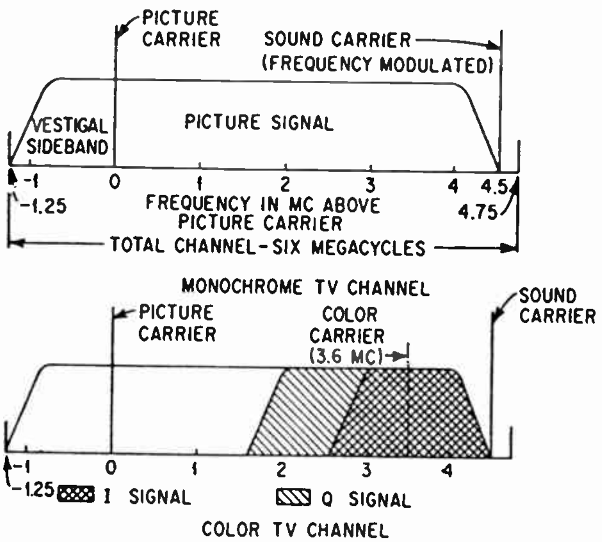


Figure 1.9-23. Diagram of television channel sing portions occupied by color and monochrome signal components.

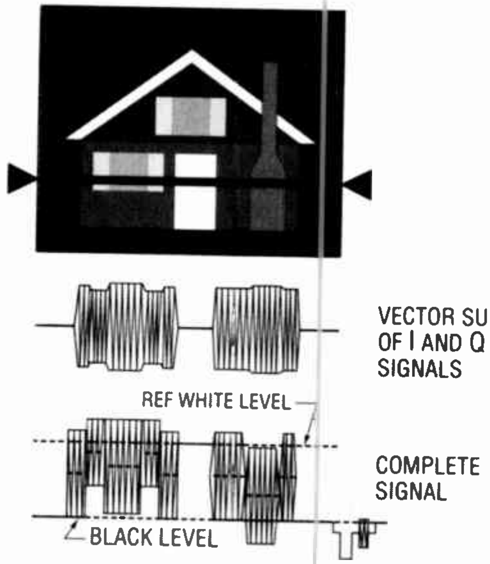


Figure 1.9-25. Typical color image.

chrome receiver, its interference effects are not objectionable because of the application of the frequency interlace principle.

The frequency interlace technique is based on two factors: a precise choice of the color subcarrier frequency and the familiar persistence-of-vision effect. If the color subcarrier is made an *odd multiple of one-half the line frequency*, its apparent polarity can be made to reverse between successive scans of the same area in the picture. Since the eye responds to the average stimulation after two or more scans, the interference effect of the color subcarrier tends to be self-canceling, owing to the periodic polarity reversals (see Figure 1.9-26).

Color Frequency Standards

The relationships among the various frequencies used in a compatible color system are illustrated in the block diagram of Figure 1.9-27. The actual frequency of the color subcarrier, which has been referred to as 3.6 MHz is specified by FCC Rules as 3.579545 MHz or exactly 455 multiplied by 1/2 the line frequency.

In broadcast practice, the frequency of the color subcarrier provides a frequency standard for operation of the entire system. A crystal oscillator at the specified frequency provides the basic control information for all other frequencies. Counter stages and multipliers derive the basic frequencies needed in the color studio.

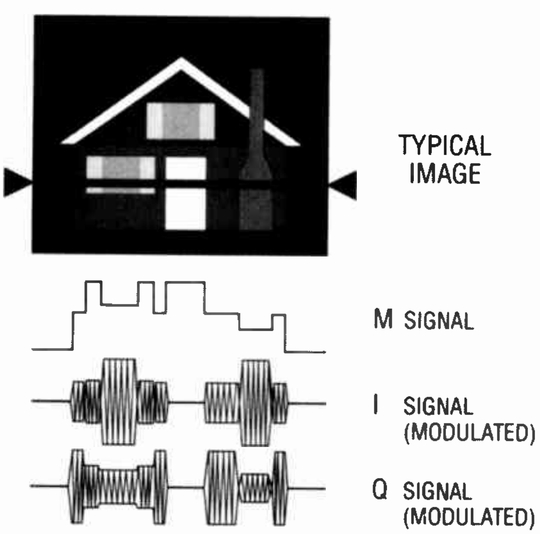


Figure 1.9-24. Typical color image and waveforms of the M signal and modulated I and Q signals.

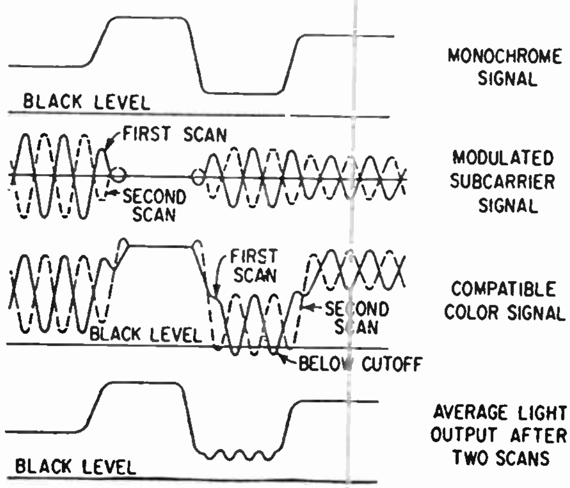


Figure 1.9-26. Waveforms showing superposition of modulated subcarrier on scanning signals, compatible color signal and effect of subcarrier on average light output.

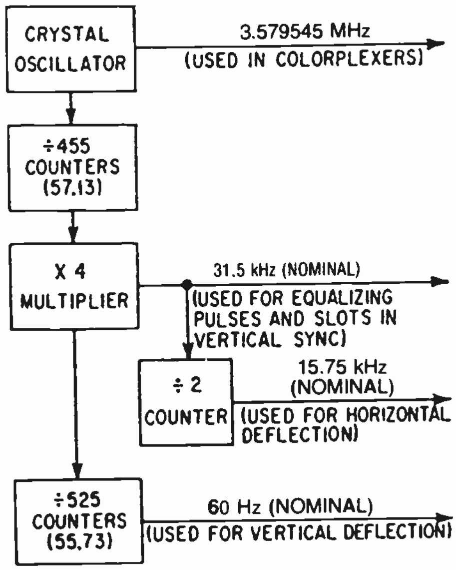


Figure 1.9-27. Block diagram showing relationship between various frequencies used in color television station.

A frequency of nominally 31.5 kHz is required for the equalizing pulses, which precede and follow each vertical sync pulse and for the serrations in the vertical sync pulse. A divide-by-2 counter controlled by the 31.5 kHz signal provides the line frequency pulses at nominally 15.75 kHz needed to control the horizontal blanking and synchronizing waveforms. Another counter chain provides the 60 Hz pulses needed for control of the vertical blanking and synchronizing circuits.

The Overall Color System

The major functions performed in transmitting and receiving color are shown in the overall block diagrams of the transmitting and receiving systems (see Figures 1.9-28 and 29).

At the transmitting end, camera output signals corresponding to the red, green and blue components of the scene being televised are passed through nonlinear amplifiers (the gamma correctors) which compensate for the nonlinearity of the picture tube elements at the receiving end. Gamma corrected signals are then matrixed to produce the luminance signal M and two chrominance signals I and Q. The filter section establishes the bandwidth of these signals. The 4.1 MHz filter for the luminance channel is shown in dotted lines because in practice this band shaping is usually achieved by the attenuation characteristics of the transmitter and the filter is not required.

The bandwidths of 1.5 and 0.5 MHz known for the I and Q channels, respectively, are nominal only; the required frequency response characteristics are described in more detail in the complete FCC signal specifications. Delay compensation is needed in the filter section in order to permit all signal components to be transmitted in time coincidence. In general, the delay time for relatively simple filter circuits varies inversely with the bandwidth. The narrower the bandwidth, the greater the delay. Consequently, a delay network or a length of delay cable must be inserted in the I channel to provide the same delay introduced by the narrower band filter in the Q channel, and still more delay must be inserted in the M channel.

In the modulator section, the I and Q signals are modulated upon two subcarriers of the same frequency but 90° apart in phase. The modulators employed are the double balanced type, allowing both the carriers and the original I and Q signals to be suppressed, leaving only the sidebands. Some sort of keying circuit must be provided to produce the color synchronizing bursts during the horizontal blanking intervals. To comply with the FCC signal specifications, the phase of the burst should be 57° ahead of the I component (which leads the Q component by 90°). This phase position was chosen mainly because it permits certain simplifications in receiver designs. Timing information for keying in the burst can be obtained from a *burst flag generator*, which in early systems was a simple arrangement of multivibrators controlled by horizontal and vertical drive pulses.

In the mixer section, the M signal, the two subcarriers modulated by I and Q chrominance signals, and the color synchronizing bursts are all added together. Provision is also made for the addition of standard synchronizing pulses, so that the output of the mixer section is a complete color television signal containing both picture and synchronizing information. This signal can then be put on-the-air by means of a standard television transmitter, which must be modified only to the extent necessary to assure performance within the reduced tolerance limits required by the color signal. (Since the color signal places more information in the channel than a black-and-white signal, the requirements for frequency response, amplitude linearity and uniformity of delay time are stricter).

The Color Receiving System

In a compatible color receiver, the antenna, RF tuner, IF strip and second detector serve the same functions as the corresponding components of a black-and-white receiver except that the tolerance limits on performance are somewhat tighter.

The signal from the second detector is utilized in four circuit branches. One circuit branch directs the complete signal toward the color picture tube, where it is used to control luminance by being applied to all kinescope guns in equal proportions. In the second circuit branch, a band-pass filter separates the high-frequency components of the signal (roughly 2.0 to 4.1 MHz) consisting mainly of the two-phase modulated subcarrier signal. This signal is applied to a pair of modulators which operate as synchronous detectors to recover the original I and Q signals. It should be noted that those frequency components of the luminance signal falling between about 2 and 4.1 MHz are also applied to the modulators and are heterodyned down

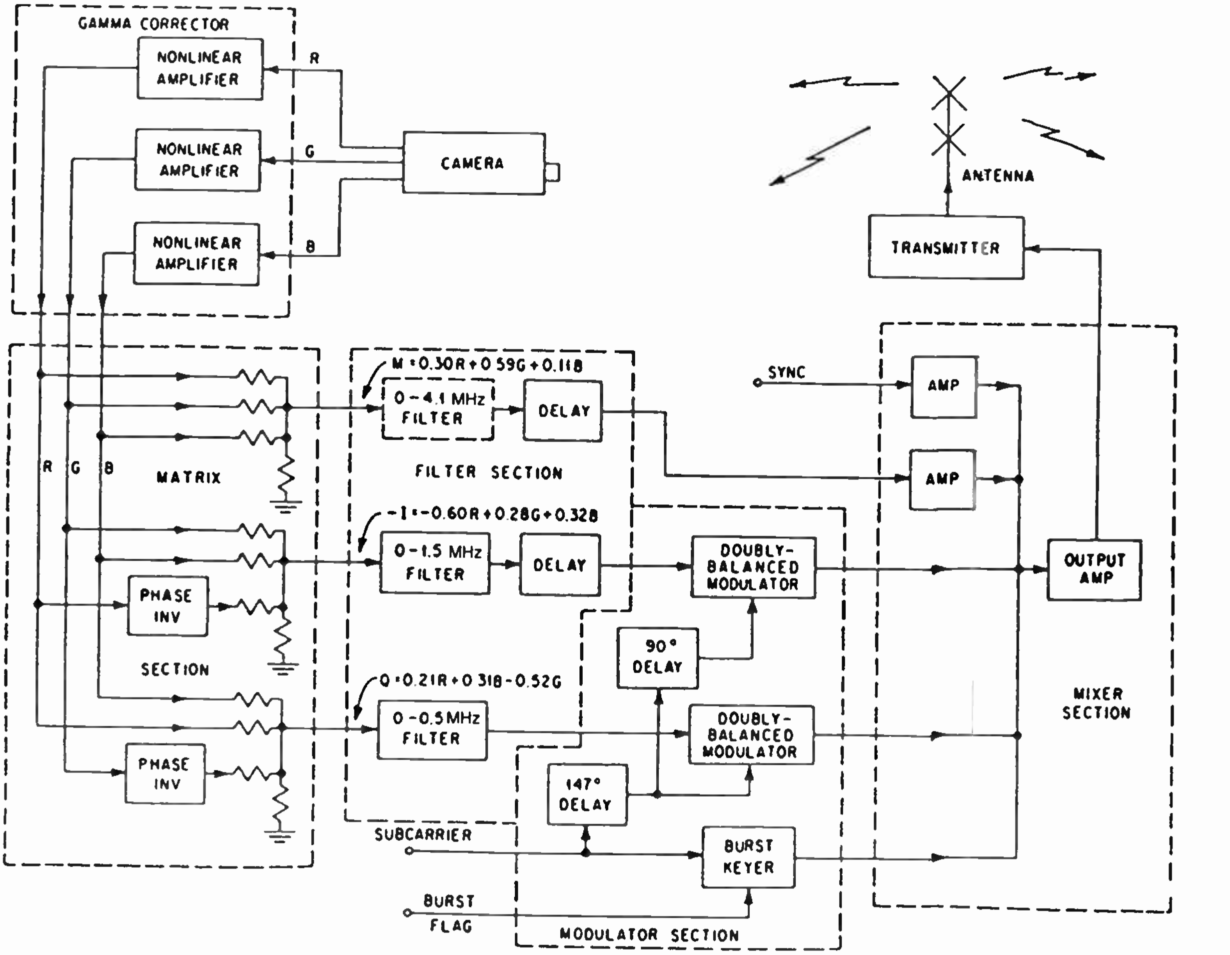


Figure 1.9-28. Block diagram showing major functions of color-transmitting system.

to lower frequencies. These frequency components do not cause objectionable interference, however, because they are frequency interlaced and tend to cancel out through persistence of vision.

The remaining two circuit branches at the output of the second detector make use of the timing or synchronizing information in the signal. A conventional sync separator is used to produce the pulses needed to control the horizontal- and vertical-deflection circuits which are also conventional. The high-voltage supply for the kinescope can be obtained either from a flyback supply associated with the horizontal deflection circuit or from an independent RF power supply. Many color CRTs require convergence signals to enable the scanning beams to coincide at the screen in all parts of the picture area; the waveforms required for this purpose are readily derived from the deflection circuits.

The final branch at the output of the second detector is the burst gate, which is turned on only for a brief interval following each horizontal sync pulse by means of a keying pulse. This pulse may be derived from a multivibrator controlled by sync pulses, as illustrated in Figure 1.9-29, or it may be derived from the flyback pulse produced by the horizontal output stage. The separated bursts are amplified and compared with the output of a local oscillator in a phase detector. If there is a phase difference between the local signal and the bursts, an error voltage is developed by the phase detector. This error voltage restores the oscillator to the correct phase by means of a varactor connected in parallel with the tuned circuit of the oscillator. This automatic frequency control circuit keeps the receiver oscillator in synchronism with the master subcarrier oscillator at the transmitter. The output of the oscillator provides the reference carriers for the two synchronous detectors; a 90° phase shifter is necessary to delay the phase of the Q modulator by 90° relative to the I modulator.

There is a *filter section* in a color receiver that is rather similar to the filter section of the transmitting equipment. The M, I and Q signals must all be passed through filters in order to separate the desired signals from other frequency components which, if unimpeded, might cause spurious effects. The I and Q signals are passed through filters of nominally 1.5 and 0.5 MHz bandwidth, respectively, just as at the trans-

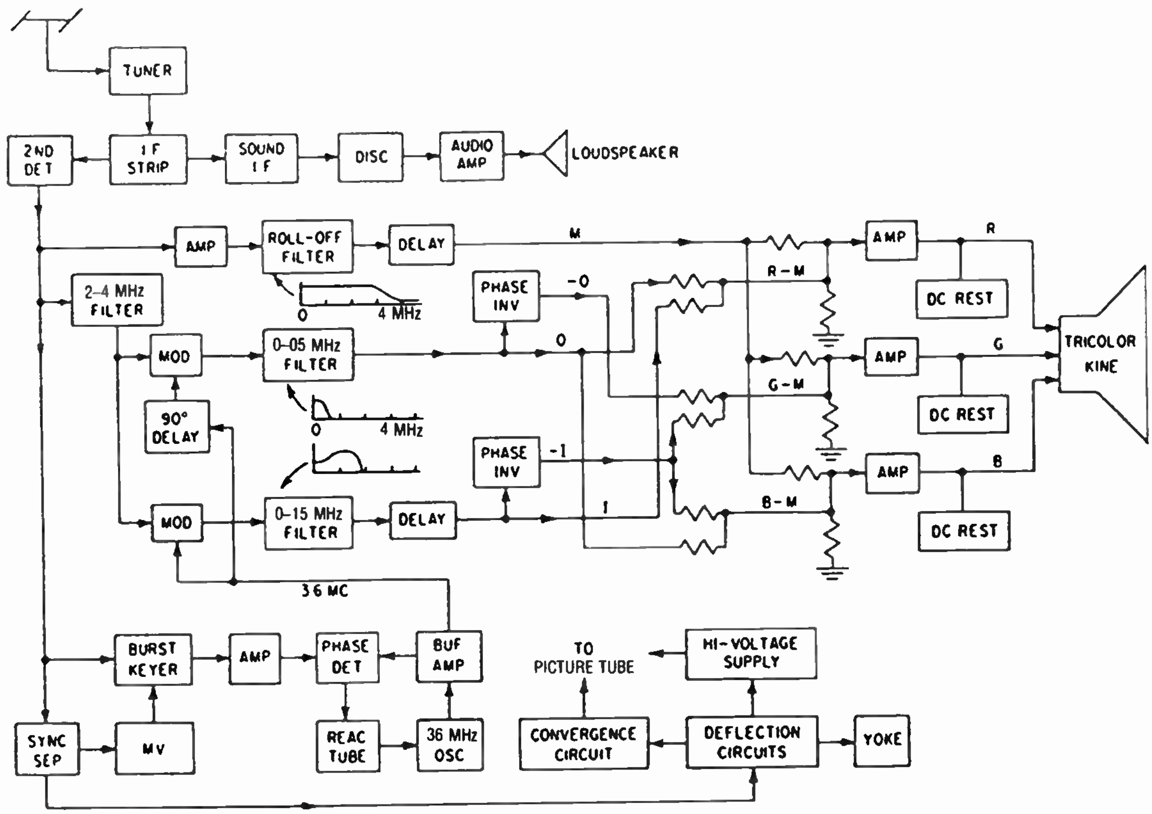


Figure 1.9-29. Block diagram showing major functions of color-receiving systems.

mitting end. A step type characteristic is theoretically required for the I filter, as indicated in Figure 1.9-23, to compensate for the loss of one sideband for all frequency components above about 0.5 MHz. Actually, this requirement is ignored in many practical receiver designs, resulting in only a slight loss in sharpness in the I channel. A roll-off filter is desirable in the M channel to attenuate the subcarrier signal before it reaches the CRT. The subcarrier would tend to dilute the colors on the screen if it were permitted to appear on the CRT grids at full amplitude. Delay networks are needed to compensate for the different inherent delays of the three filters, as explained previously.

Following the filter section in the receiver there is a matrix section in which the M, I and Q signals are cross-mixed to recreate the original R, G and B signals. The R, G and B signals at the receiver are not identical with those at the transmitter because the higher frequency components are mixed and are common to all three channels. This mixing is justifiable because the eye cannot perceive the fine detail (conveyed by the high-frequency components) in color. There are many possible types of matrixing circuits. The resistance mixers shown provide one simple and reliable approach. For ease of analysis, the matrix operations at the receiver can be considered in two stages. The I and Q signals are first cross-mixed to produce R-M, G-M and B-M signals (note that *negative* I and Q signals are required in some cases), which are, in turn, added to M to produce R, G and B.

In the output section of the receiver, the signals are amplified to the level necessary to drive the kinescope and the dc component is restored. The image which appears on the color CRT screen is a high quality full color image of the scene before the color camera.

It should be made clear that the block diagram of Figure 1.9-29 is intended only to illustrate the principles used in color receivers and does not represent any specific model now on the market. Design engineers of color receivers have shown great ingenuity in simplifying circuits, in combining functions, and in devising subtle variations in the basic process which have made possible significant cost reductions while maintaining excellent picture fidelity.

COLOR FIDELITY

Color fidelity as used herein, is the property of a color television system to reproduce colors, which are realistic and pleasing to the average viewer.

Although perhaps not apparent at first, color fidelity is analogous to high fidelity as applied to sound reproduction. Just as a high fidelity audio system faithfully reproduces sounds reaching the microphone, the color television system is capable of faithfully reproducing colors as seen by the television camera. In fact, the color television system is capable of reproducing colors more accurately than techniques presently used in color printing and color photography.

Tests have shown, however, that color television pictures are generally more pleasing to viewers when deliberate modifications are made in the reproduced colors to compensate for the surroundings in which they are reproduced. The situation is similar to that experienced in the art of sound reproduction in the case of a symphony orchestra recorded at high sound levels in a large hall and reproduced at lower sound levels in a small room. In this case, a more pleasing effect is obtained if the ear's new environment is taken into consideration and the reproduction modified accordingly. Similarly, in color television, the changed environment of the eye must be considered and the reproduced colors modified accordingly.

Color fidelity, therefore, is a term used to indicate a color reproduction, which pleases viewers aesthetically and persuades them that they are viewing a faithful reproduction of the original colors in the scene being televised.[1]

The following describes possible distortions in the color system and their effect on the picture and prescribes amounts or degrees of distortion that can be tolerated without adverse effects on picture quality.

Color System Analysis

Individual elements or areas of the complete color system are discussed in the following paragraphs with the aid of the diagrams shown in Figures 1.9-30 through 34.

Figure 1.9-30 is a theoretical color system in that it assumes linear camera pickup devices and a linear picture tube interconnected by a distortionless wire system. The only distortion that can result from this system is a flaw in colorimetry.

Figure 1.9-31 introduces linearity correctors to compensate for color errors produced by nonlinearities in the transducers.

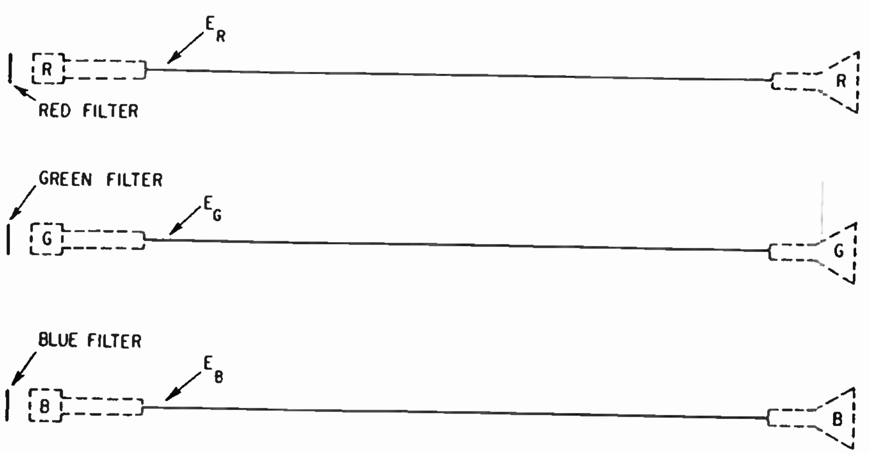


Figure 1.9-30. Diagram of theoretical color system showing linear RGB pickup elements and picture tubes interconnected by wire.

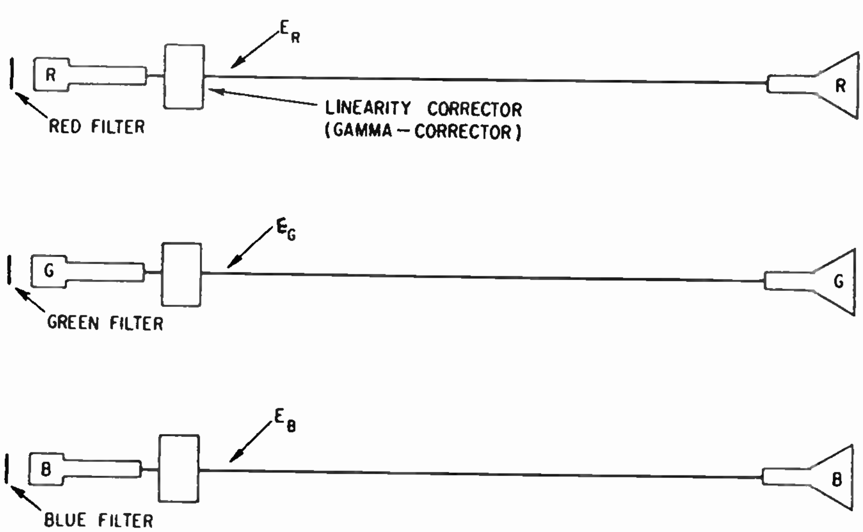


Figure 1.9-31. The basic color system shown with necessary linearity correctors to compensate for color errors introduced by the nonlinear transducers.

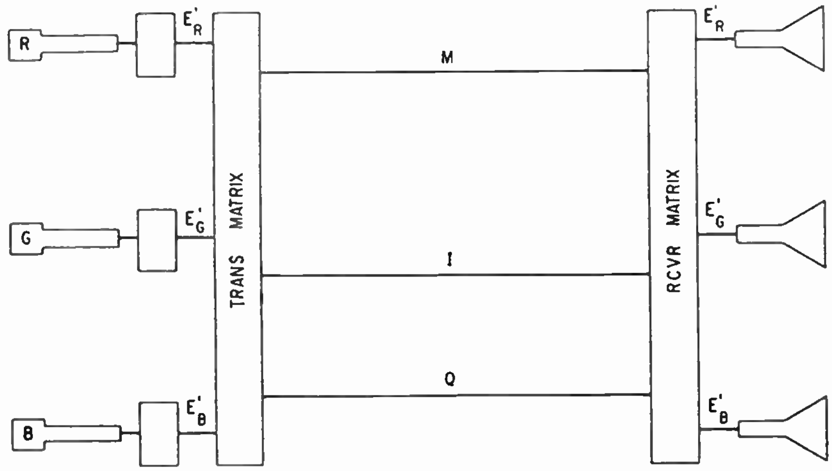


Figure 1.9-32. Diagram showing transmitter and receiver matrix functions in the color system.

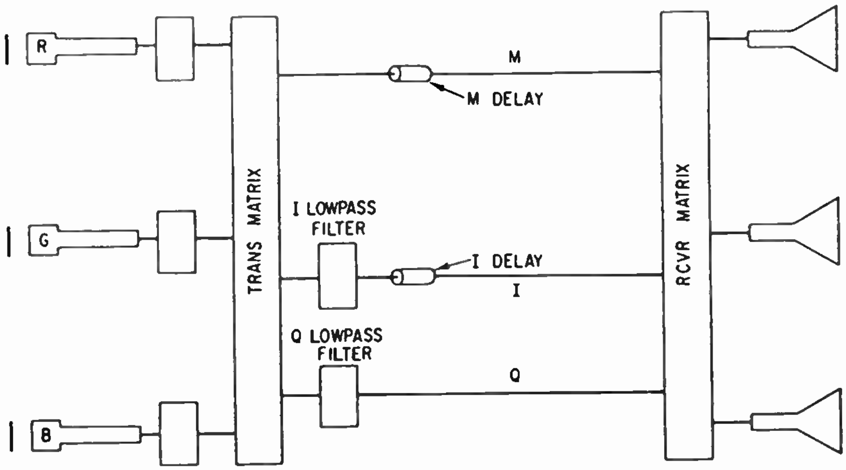


Figure 1.9-33. Basic color system with band limiting and delay compensation.

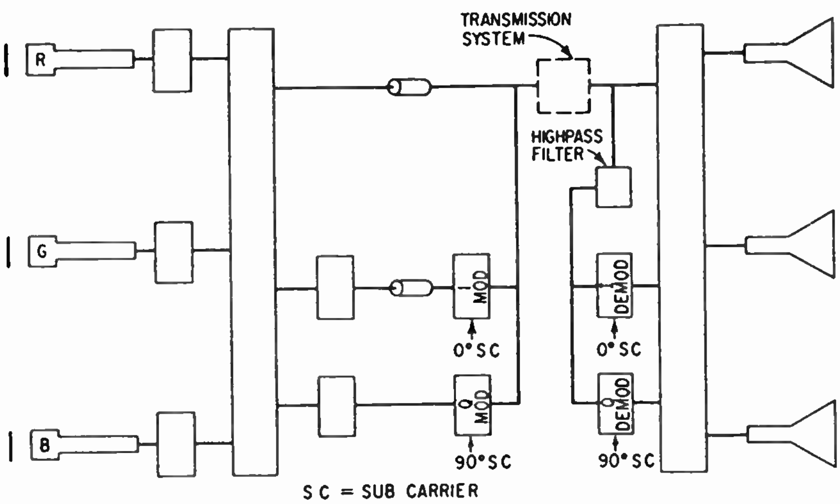


Figure 1.9-34. Basic color system showing all major elements, including the transmission system.

Figure 1.9-32, 33, and 34 successively introduce the complexities of matrixing, band limiting, delay compensation and the transmission system (shown dotted in Figure 1.9-34). These diagrams, each representing a possible color system, introduce techniques used in compatible color television and permit the study of color distortions peculiar to each technique.

The systems diagrammed in Figures 1.9-30 and 31 are described under *Possible Distortions in Transducers*, and those in Figures 1.9-32, 33 and 34 under *Possible Distortions in Encoding and Decoding Processes*. The system shown in Figure 1.9-34 is discussed under *Distortions in the Transmission System*.

Characteristics of the Eye

To appreciate fully the significance of color fidelity, it is helpful to consider some of the characteristics of the eye that are associated with color perception and to analyze such terms as color adaptation, reference white and primary colors and determine their relationship to a color television system.

Color Adaptation

One amazing characteristic of the eye is the phenomenon known as *color adaptation*. It is this adaptation which enables one to describe accurately the color of an object under white light while viewing in nonwhite light. That is to say, recognition of color is surprisingly independent of the illumination under which an object is viewed. For example, if sunlight at high noon on a cloudless day is taken as white light, then, by comparison, the illumination from a typical 100 W incandescent bulb is very yellow light. Yet it is known that an object viewed under sunlight looks very little if any different when viewed under incandescent light. Moreover, it is obvious to the observer, after a very few minutes in a room illuminated with incandescent lights, that the light is not yellow at all; it is really white.

It is apparent, then, that the color seen by an observer is dependent upon the illumination to which that observer has been exposed for the past several minutes. This ambient illumination will have a marked effect on the choice of color to be called white.

This phenomenon can cause a loss of color fidelity under certain conditions. Consider, for example, a theoretically perfect color system with the camera viewing an outdoor scene under a midday sun while the reproduced picture is being viewed in a semidarkened room, with what little light is in the room also being derived from the midday sun. Under these conditions, the ambient illuminations at both camera and receiver are identical, so a person standing alongside the camera and one viewing the receiver would both see the same colors. Now, if a change in the weather at the camera location should cause a cloud to cover the sun, the ambient illumination at the camera location would shift toward a bluer color. This shift would not disturb the viewer standing alongside the camera, because the observer's eyes, bathed in the new ambient light, would rapidly adapt to the new viewing conditions and they would perceive the scene as being unchanged.

The person viewing the receiver would not be so fortunate. Assuming that they are far enough away that this same cloud would not affect their ambient, they would observe that everything on the screen had suddenly and inexplicably taken on a bluish cast, which would certainly be most disturbing.

Such errors in color fidelity can be corrected by making the camera imitate the human eye in adaptation. The eye adapts to changes in ambient illumination by changing its sensitivity to a certain color. For example, if a light source changes from white to blue-white (as in the example), the eye reduces its blue sensitivity until the light again appears to be white to the observer. Likewise, a camera operator can correct for the same situation by decreasing the gain of the blue channel of the camera or by attenuating the light reaching the blue camera tube. In this way, the camera is made to color adapt, and the reproduced picture on a receiver loses its bluish cast.

Reference White

Although color adaptation can generate a problem such as the one just described, it also simplifies certain requirements. Specifically, it eases the requirement that white be transmitted as a definite, absolute color, for there clearly can be no absolute white when almost any color can be made to appear subjectively white by making it the color of the ambient illumination to which an observer's eye has adapted.

In color television, this characteristic is taken advantage of in the following manner: A surface in the studio which is known by common experience to be white, for example, the Electronic Industries Alliance (EIA) Gray-Scale Chart of a piece of Neutracor white paper, is selected to be reproduced as white on a home receiver. The relative sensitivities of the three-color channels of the camera are then adjusted so that the camera adapts to this white regardless of the studio illumination. The home receiver can then be adjusted to reproduce the surface as any white which the home viewer prefers, depending upon his surroundings.

It has already been mentioned that the eye adapts readily to the illumination that surrounds conditions of an overcast day. This representative standard illumination has been adopted internationally as a base for the specification of the color of objects when they are viewed outdoors. This standard (Illuminant C) has been chosen to be the "standard-viewing-white" of the receiver.

The change in reference white between studio and home will inevitably produce errors in all reproduced colors, but the errors are small and, more important, tend to be subjectively self-correcting, so that any given object will produce the same color sensation whether viewed in relation to the studio reference white or the home reference white.

Consequently, a viewer may become familiar with an object such as a sponsor's packaged product and will recognize it on the television screen, under the fluorescent lighting of the supermarket, or under the incandescent lighting of the home and, furthermore, will note little difference in the colorimetric values of the package under the three conditions, even though the absolute colorimetric values would be appreciably different in the three situations.

Primary Colors

Of all the characteristics of the eye, there is perhaps none more fundamental to practical color television than that characteristic which allows viewers to choose certain colors called primary colors, and from these synthesize almost any other desired color by adding together the proper proportions of the primary colors. If it were noted for this characteristic, each hue in a color system would have to be transmitted over a separate channel; such a system would be too awkward to be practical. Because of the eye's acceptance of synthesized colors, it is possible to provide excellent color rendition by transmitting only the three primary colors in their proper proportions.

Possible Errors in Transducers

The block diagram of Figure 1.9-30 shows a fundamental color television system using red, green and blue primaries and three independent transmission channels. The camera tubes and kinescopes are shown dotted to indicate that any inherent nonlinearities in these devices are to be disregarded, for the moment, in order to simplify the discussion of the colorimetry of the system.

The general plan is a system, such as Figure 1.9-30, to provide the three picture tubes with red, green and blue phosphors, respectively, and to allow the corresponding camera tubes to view the scene through an appropriate set of red, green and blue filters. If a phosphor and a filter have the same dominant wavelength, that is, if they appear to the eye to be the same color, it might be mistakenly supposed that they would be colorimetrically suited to be used as a filter and phosphor set for the channel handling that color. Actually, the basis for choosing filters and phosphors is much more complex and is based on the shape of the response curve of the filter, plotted against wavelength, and the shape of the light output curve of the phosphor, also plotted against wavelength. The following paragraphs will discuss briefly a technique which might be used to determine the required relationship between the phosphor curves and the filter curves.

The color characteristics of the phosphor are generally less easily changed than are filter characteristics; for this reason characteristics of phosphors are taken as the starting point, and characteristics of the filters are determined from them. A laboratory setup which could be used to determine these characteristics is shown in Figure 1.9-35. In this figure, an observer (who must have normal vision) is simultaneously viewing two adjacent areas, one of which is illuminated by a source of single-wavelength light which can select any wavelength in the visible spectrum, the other of which is illuminated by a red picture tube, a green picture tube and a blue picture tube. The phosphors of these CRTs are the phosphors which are to be used in the color system. Starting at the red end of the spectrum, a single-wavelength red is selected to illuminate the left-hand area, and the light from each of the three phosphors is varied until a color match is obtained between the left-hand and right-hand areas. The respective amounts of red, green and blue lights needed to accomplish this match are recorded. Then another wavelength is chosen, the kinescope outputs varied to produce a match, and the new amounts of red, green and blue needed for a match are recorded. Similarly, points are obtained throughout the entire spectrum, and a graph is plotted showing the various required outputs versus wavelength. The shapes of these three curves (one for red, one for green and one for blue) are the required shapes for the three camera-filter response curves. The resulting curves would in general resemble Figure 1.9-36.

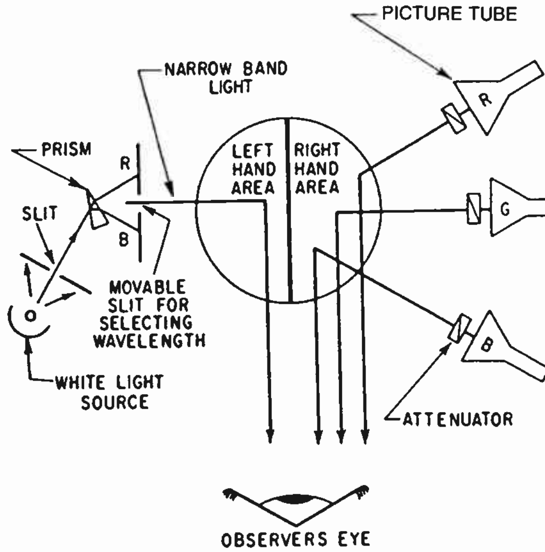


Figure 1.9-35. Diagram showing laboratory setup arranged to compare narrow-band light source and R, G and B light produced by picture tubes to determine proper camera-filter color characteristics.

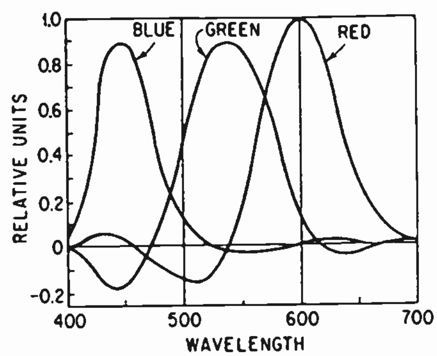


Figure 1.9-36. Curves showing relative quantities in camera output required to produce correct picture tube colors over the visible spectrum.

(To simplify the above discussion it was assumed that the camera pickup devices responded equally well to all wavelengths. In practice, tubes show higher output at certain wavelengths than at others. The filter-response curves derived by the technique would have to be modified so that the combined response of filter and camera would be correct.)

Certain practical difficulties could result in errors in the previous procedure. For example, if the observer had any deviations from normality in color vision characteristics (as most people do), these deviations would result in nonstandard matches and, hence, improper camera filter characteristics. Also, if the phosphors were contaminated in any way during their manufacturing process (as most phosphors are, at least to some small degree), the resulting phosphor characteristics would not be the proper ones and hence would give rise to improper camera filter characteristics. The observer errors can be normalized out by standard colorimetric procedures, but phosphor errors represent a basic error which may possibly be present not only in the previous experiment but also in varying degrees in a large number of receivers. Quality control of phosphor manufacture is sufficiently good, however, to make the net effect unnoticeable in home receivers.

A striking practical difficulty would also arise regardless of observer or phosphor errors. For most wavelengths, no combination of red, green and blue picture tube outputs could be found which would produce a match. In order to obtain a match at these wavelengths, it would be necessary to move one or two of the CRTs over to the other side so that they could add their light to the single-wavelength light being matched. This procedure can be described mathematically, for graphing purposes, by saying that adding light to the left-hand area is the same as subtracting light from the right-hand area. Therefore, the amount of light added on the left would be considered as a negative quantity and would result in a point below the axis on the graph. Since this condition would be found to exist for several successive wavelengths, the resulting graph would show one or more minor lobes below the axis. These are called *negative lobes*.

These negative lobes represent a need for filters with negative light transmission characteristics at certain wavelengths. Simple attenuating filters cannot yield such a characteristic, much more elaborate means would be required.

It is theoretically possible to achieve these negative lobes with added camera complexity but it has been shown that excellent color fidelity can be obtained by ignoring the negative lobes and using filters, which yield the positive lobes only. Positive lobe processes such as color photography have gained wide acceptance for years. Masking techniques which employ electrical matrixing have been introduced which can modify the spectrum characteristics of a color camera. These techniques can be used to help compensate for deficiencies in the color fidelity such as the lack of negative lobes.

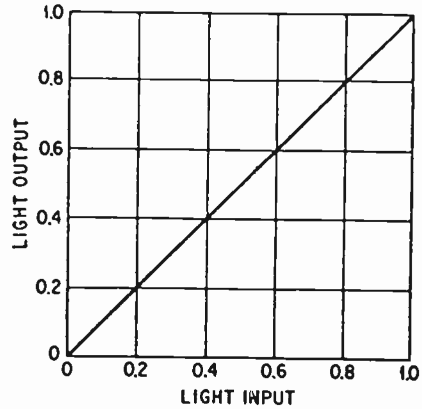


Figure 1.9-37. Curve showing light transfer characteristics of a perfectly transparent piece of window glass.

Transfer Characteristics

A piece of window glass is perhaps the nearest approach to a perfect video system. For a piece of glass, the light output (to the viewer) is essentially identical with the light output (from the scene). This fact is shown graphically in Figure 1.9-37. This plot could be called the *transfer characteristic* of a piece of glass, since it describes the way that light is transferred through the system.

If the window glass is replaced by a neutral density filter, which attenuates light 3-to-1, the transfer characteristic will then be given by Figure 1.9-37. The difference between Figure 1.9-37 and 38 can be described by these simple relationships:

For the glass:

$$\text{Light output} = \text{light input}$$

For the neutral density filter:

$$\text{Light output} = k \times \text{light input}$$

where: $k = 1/3$ in this case.

Both systems are linear—doubling the light input of either will double its light output, tripling input will triple output, etc. A nonlinear system does not exhibit this simple proportionality. For example, consider a system described by

$$\text{Light output} = k \times (\text{light input})^2$$

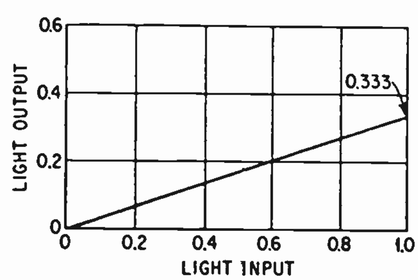


Figure 1.9-38. Curve showing transfer characteristic of a neutral density filter with 3-to-1 light attenuation.

Doubling the input to this system will quadruple its output, a threefold increase in input will result in a ninefold increase in output, etc. The transfer characteristic for this type of system is shown in Figure 1.9-39. Note that the characteristic is definitely nonlinear; that is, it is not a straight line as were Figures 1.9-37 and 38.

In television and photography, nonlinearity is more common than linearity. For example, an ordinary picture tube is a nonlinear device, having a transfer characteristic which can be approximated by the expression:

$$\text{Light output} = k(\text{voltage input})^{2.2}$$

Camera tubes can be linear or nonlinear devices. For example, the characteristic of a vidicon is approximately

$$\text{Current input} = k(\text{light input})^{0.65}$$

The general expression for nonlinear transfer characteristic can be given approximately as

$$\text{Output} = k(\text{input})^{\gamma}$$

where the exponent is the Greek letter gamma.

Graphical Displays of Transfer Characteristics

The first reaction of any person asked to display two variables (like light input and light output) on a set of XY coordinates is to divide X and Y coordinates into equal increments and plots the variables in this manner. A typical result of such a plot has already been described (see Figures 1.9-37 and 38). Such a plot has the advantage of showing at a glance the linearity of the device described by the variables. If the plot is a straight line, the device is linear; if curved, the device is nonlinear. Moreover, the slope of the line describes the attenuation (or gain) of the device. If the slope is unity (which occurs when the plot makes a 45° angle with the X axis), there is no attenuation. It is a very good piece of glass. For the neutral density filter previously described, which has the equation

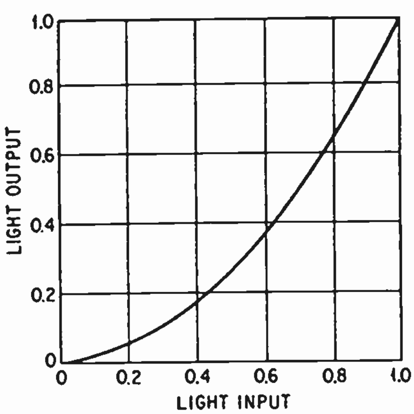


Figure 1.9-39. Curve showing a nonlinear transfer characteristic.

(light output) = 1/3 (light input), the line has a slope of one-third (see Figure 1.9-38).

Such are the advantages of plotting transfer characteristics with equal increment divisions of the X and Y axis. However, other advantages (very important ones) can be obtained by dividing up the X and Y coordinates logarithmically. Such a plot is called a *log-log* plot.

Consider a system which has a transfer characteristic given by $L_0 = (L_{in})^{2.2}$. If this equation is plotted on axes which are divided logarithmically, the resulting plot is the same as though the logarithm of both sides of the equation were plotted on equal increment axes. Taking the logarithm of both sides, we obtain

$$\log L_0 = \log (L_{in})^{2.2}$$

Since $\log (L_{in})^{2.2}$ is the same as $2.2 \log (L_{in})$, then

$$\log L_0 = 2.2 \log L_{in}$$

Comparing the form of this equation with an earlier equation, light output = 1/3 light input, the attenuation, 1/3, was the slope of the earlier equation, so 2.2, the exponent, is the slope of the latter equation. The use of logarithmically divided coordinates yields a plot in which the exponent is given by the slope of the line. Therefore, this plot will show at a glance the magnitude of the exponent and will also show whether or not the exponent of the system is constant for all light levels. It also is advantageous in showing the effects of stray light.

Figures 1.9-40(a) and 40(b) compare the two types of plotting for three types of transfer characteristics.

The Effect of a Nonlinear Transfer Characteristic on Color Signals

In monochrome television, some degree of nonlinearity can be tolerated, but such is not the case for a color television system. It can be shown that a system exponent different from unity must inevitably cause a loss of color fidelity. For an example, consider a situation in which signals are being applied through linear amplifiers to the red and green guns of a perfectly linear (theoretical) picture tube. The green amplifier is receiving 1.0 V; the red amplifier, 0.5 V. If everything is perfectly linear, the proportions of the light output should be 1.0G + 0.5R = greenish yellow. However, if the CRT has an exponent of 2.0, the light output will be $(1.0)^2G + (0.5)^2R = 1.0G + 0.25R$ = greenish yellow with an excess of green.

From the previous specific case, it may be correctly inferred that in general, a system exponent greater than 1 will cause all hues made of the combination of two or more primaries to shift toward the larger or largest primary of the combination. Conversely, a system exponent less than 1 will shift all hues away from the largest primary of the combination.

In the previous example, an exponent of 0.5 would yield $(1.0)^{0.5}G + (0.5)^{0.5}R = 1.0G + 0.70^{7}R$ = a greenish yellow which is a shade off a pure yellow.

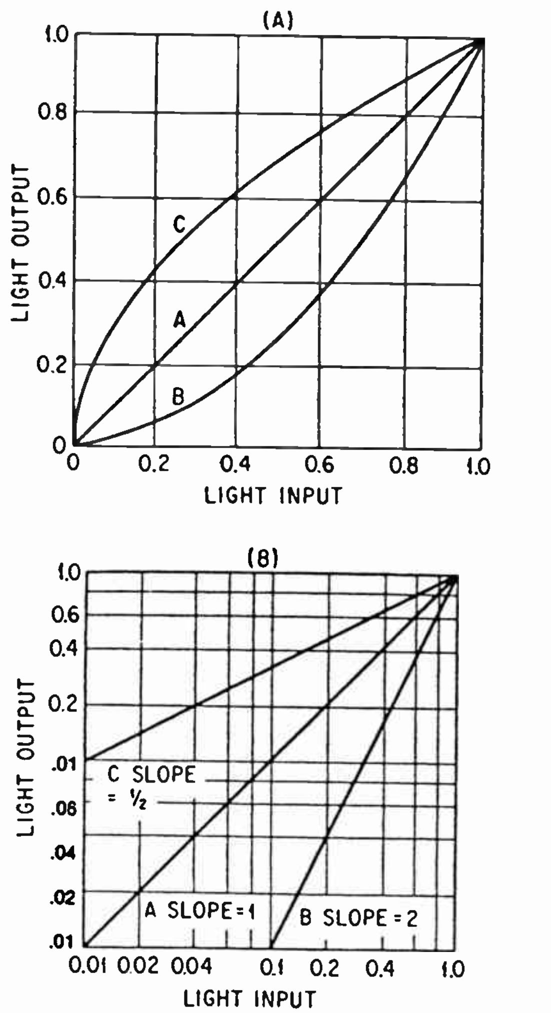


Figure 1.9-40. Graphs showing the curves obtained by plotting A, B and C types of transfer characteristics on linear coordinated (a) and on log-log coordinates (b).

In addition, the reader can correctly conclude that white or gray areas, in which all the primaries are equal, will not be shifted in hue by a nonunity exponent.

Effect of Differing Exponents in Each Channel

The preceding discussion assumed that all three channels (in Figure 1.9-30) have the same exponent, whether in unity or not. In practical systems, however, there is always the possibility that the exponents of the channels may differ from one another. This situation will produce intolerable color errors if the differences become even moderately large. In general, the requirements for tracking among the light transfer characteristics of the individual channels are even more stringent than the requirement for unity exponent.

Figures 1.9-41(a), 41(b), 41(c) and 41(d) show the effects of unequal exponents in the three channels. In all four figures, the red and blue exponents are taken as unity; in Figures 1.9-41(a) and 41(b) the green exponent is taken as less than 1, and in Figures 1.9-41(c) and 41(d), as greater than 1. In Figure 1.9-41(a), the transfer characteristics are shown for the system adjusted to produce peak white properly. It can be seen that the bowed characteristic of the green channel will cause all whites of less than peak value to have too much green. A gray-scale step tablet before the camera would be reproduced properly only at peak white; the gray steps would all have a greenish tinge. Relative channel gains could be readjusted to reproduce one of the gray steps properly (Figure 1.9-40(b)), but then all highlight steps would be purplish while lowlight steps would still be greenish.

A green-channel exponent greater than unity would reverse those results (Figure 1.9-41(c) and 41(d)). With gains adjusted to reproduce peak white properly (Figure 1.9-41(c)) lowlights would be purplish; with gains readjusted to provide proper reproduction for one of the lower steps (Figure 1.9-41(d)) highlights would be green and lowlights purple.

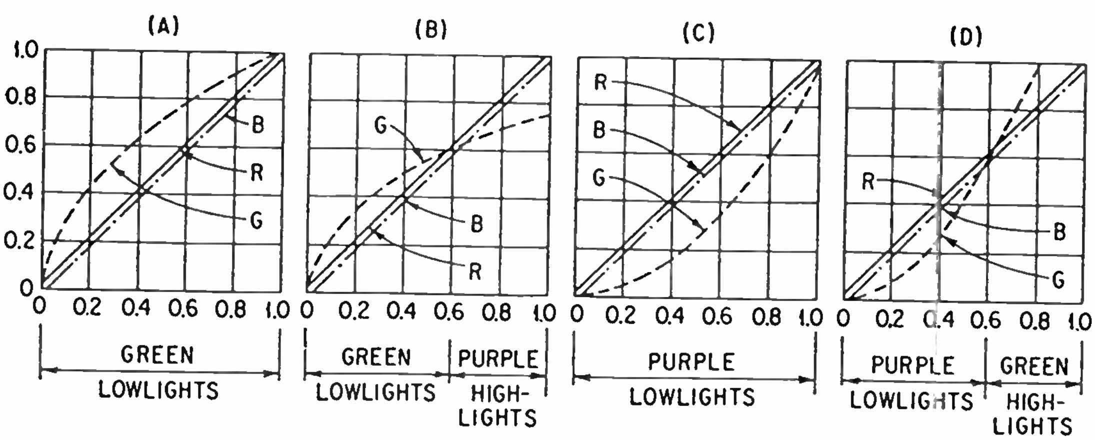


Figure 1.9-41. Linear plots showing graphically the effect of unequal exponents in the R, G and B channels. In all four graphs the R and B exponents are taken as unity. In (a) and (b) the green exponent is taken as less than 1, and in (c) and (d), as greater than 1.

The Effect of Stray Light

If a picture tube is viewed in a lighted room, there will always be some illumination on the faceplate. Therefore, the eye will always receive some *light output* from the picture tube, regardless of the magnitude of the signal input voltage. Under this condition, a true black is impossible to obtain.

This condition is reflected in the transfer characteristic of the system. If, for example, the stray light were 5% of the peak highlight brightness of the picture, a linear plot of light output versus light input would have the entire transfer characteristic shifted upward by 5%. However, the most interesting change is found in the log-log plot, where, as seen in Figure 1.9-42, the stray light causes a change in the slope in the lowlight regions. Since the slope is equal to the exponent, this change shows that stray light causes an effective exponent error in the lowlight regions of the picture and hence will cause color fidelity errors, which will be most marked in lowlight regions.

These errors will be noted by an observer as improper hues and saturations, with the saturation errors (a washing out of the more saturated lowlight areas) being the more objectionable to a viewer.

Stray light is not the only cause of errors of this type. Similar effects will be noted whenever the CRT bias (brightness) is set too high, if the camera pedestal is set too high, or if stray light enters the camera (whether through lens flare or any other source). In general, any condition which prevents the light output of the system from becoming zero when the light input is zero will cause errors similar to those caused by stray light.

Linearizing a System

It can be shown that a system using a vidicon with an exponent of 0.675 to drive a CRT with an exponent of 2.2 will have an overall exponent given by the product $0.65 \times 2.2 = 1.43$, assuming that all devices in the system are linear. In general, the overall exponent of a system is the product of the exponents of the cascaded elements.

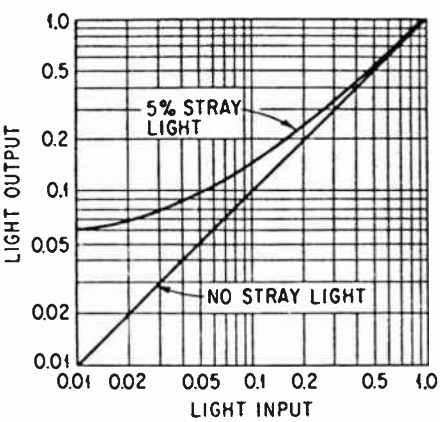


Figure 1.9-42. Log-log plot of system with stray light, illustrating change of slope in the low-light regions.

This knowledge provides an excellent tool for linearizing a system. For example, a system with an overall exponent of 1.43 could be linearized by inserting somewhere (in a video path) an amplifier having an exponent of 1/1.43 (= 0.7) so that the product becomes unity: $1.43 \times 1/1/43 = 1$.

In Figure 1.9-31, a nonlinear amplifier, or gamma corrector, is shown inserted in each of the three paths.

Possible Encoding and Decoding Distortions

The second of the two systems discussed in the preceding section bordered on being a practical system but still required three independent 4 MHz channels. A fortunate characteristic of the human eye (the inability to see colored fine detail) allows us to modify this requirement to one 4 MHz channel for monochrome fine detail and two much narrower channels for color information. Before this modification can be made, the red, green and blue signals must be combined to form three other signals, usually called M, I and Q, such that the M signal alone requires a 4 MHz channel, and the I and Q channels, which contain the color information, are confined to narrower channels. This rearrangement of red, green and blue to form M, I and Q is called matrixing and was described in the previous section. A system which uses a matrix is block diagramed in Figure 1.9-32. The illustration also shows that to recover the original red, green and blue signals at the receiving end, a rearranging device is needed. This device is usually called the *receiver matrix.*

Matrixing alone offers no advantage unless steps are taken actually to limit the I-signal and Q-signal channels to the narrow bandwidths allowed. Figure 1.9-33 shows a system employing such band shaping. The band shaping filters themselves always introduce delay, which must be compensated for by placing delay lines in the wider band channels, as shown in the diagram.

To put both color and monochrome information in the spectrum space normally occupied by monochrome only requires that the color information overlap the monochrome. This overlap can be allowed for both I and Q signals, without incurring visible crosstalk, if frequency interlace and two-phase modulation are employed. A system using these techniques, which were described, in the section on *Electronic Aspects of Compatible Color Television,* is block diagramed in Figure 1.9-34.

Possible Errors in the Matrixing Process

The entire matrixing process can be summed up in two sets of equations, the first describing how the transmitter matrix takes in red, green and blue and turns out M, I and Q:

$$M = 0.30R + 0.59G + 0.11B$$
$$I = 0.60R - 0.28G - 0.32B$$
$$Q = 0.21R - 0.52G + 0.31B$$

and the second describing how the receiver matrix takes in M, I and Q and recreates red, green and blue:

$$R = 0.94I + 0.62Q + M$$
$$G = 0.27I + 0.67Q + M$$
$$B = 1.11I + 1.7Q + M$$

Both matrices continuously compute the desired output from the given input. The coefficients in the previous six equations are usually determined by precision resistors or, in the case of negative numbers, by precision resistors and signal-inverting amplifiers. The basic error that can occur, therefore, is a change in a resistor value or an amplifier gain, resulting in a change in one or more coefficients. In general, the resulting picture error resembles cross talk among the primary colors.

More specifically, the transmitter matrix can have two distinct types of errors. The first involves the coefficients of the equation for M, the second, the coefficients for I and Q. An error in an M coefficient will brighten or darken certain areas. In a monochrome reproduction of a color signal, such an error, if small, would not be noticed; if large, it would still probably be tolerated by the average viewer. In a color reproduction, however, even a small error would be objectionable. For example, a reduction of the red coefficient from 0.3 to 0.2 would cause a human face to be reproduced with an unnatural ruddy complexion and dark red lips.

Note that the sum of the M coefficients is 1. An error in one coefficient would change this sum, so that peak white would no longer occur as 1 V. An operator could mistake this condition for a gain error and adjust either M gain or overall gain in an effort to obtain the correct peak white voltage. Changing M gain would cause errors to occur in all M coefficients; changing overall gain would put errors in all coefficients. Although such an error is rare in well engineered equipment, it is a possible source of color error, which can be compounded by misdirected attempts at correction.

Note that the sums of the Q and I coefficients are each zero, which means that when R = G = B (the condition for white or gray), Q and I both equal zero. An error in a Q or I coefficient would cause color to appear in white or gray areas and, in addition, would cause general errors in colored areas resembling cross talk among the primaries. Controls are usually provided in the Q and I matrices, called *Q white balance* and *I white balance,* respectively, which allow the operator to adjust the sum of the Q or I coefficients by changing the value of one of the coefficients. If the coefficient controlled is the one in error, then adjusting white balance restores the condition that the sum of the coefficients is zero, that is, it removes the color from white and gray objects, but it does so by giving the controlled coefficient an error which just counteracts the error of a nonadjustable coefficient, so that two coefficients are wrong instead of one. Again, such an error is rare in well engineered equipment, for the adjustable coefficient is usually the one in error. However, the possibility of an error compounded by adjustment should be kept in mind.

A far more likely cause of white balance error is an error in input level, that is, a discrepancy between the peak white levels of input red, green or blue. In such a case, an operator can still achieve white balance (Q and I = 0 for white input) but the entire system will be in error. The starting point for all investigations of the cause of white balance errors should be the levels of the red, green and blue inputs.

In the receiver matrix, only one general type of error can occur instead of two as in the case of the transmitter matrix. This type of error, a *general coefficient error,* results in cross talk among the primary colors. For example, a change in the I coefficient for the red equation from 0.94 to 0.84 would yield about a 7% reduction in the peak red output available and would also result in unwanted red light output in green or blue areas at about 3-1/2% of the green or blue level.

Gain Stability of M, I and Q Transmission Path

In the system of Figure 1.9-32, every gain device or attenuating device in the three transmission paths must maintain a constant ratio between its input and output in order to maintain the proper ratios among the levels of M, I and Q at the input to the receiver matrix. A variation in the gain of one of these paths will result in a loss of color fidelity.

For example, a reduction in M gain must obviously cause a reduction in the viewer's sensation of brightness. Not quite so obvious are the effects of I and Q gain. Since these are color signals, their amplitude would be expected to influence the sensation of saturation, but the manner of this influence is not intuitively obvious until the factors which influenced the selection of I and Q compositions are recalled. It previously was pointed out that the eye has the greatest need for color detail in the color range from orange to blue-green (cyan) and the least in the range from green to purple. Hence I, the wider band signal, conveys mainly orange and cyan information, and Q, the narrower band signal, conveys principally the greens and purples. Therefore, a reduction in I gain could be expected to reduce the saturation sensation for colors in the orange and cyan gamut, leaving the greens and purples virtually unaffected. Conversely, Q gain will influence the greens and purples without causing much change in the appearance of orange and cyan objects.

Modulation and Demodulation

The system depicted in Figure 1.9-32, which introduced bandwidth limiting of the I and Q signals in accordance with the capabilities of the eye to see colored fine detail, is a fairly practical and economical system, except for the fact that three individual transmission channels are employed. To have a compatible system, however, these three channels must be reduced to one through some multiplexing technique. The technique used has already been described, and a system employing this technique is block diagrammed in Figure 1.9-34.

Possible Errors in Modulation

Burst Phase Error

Perhaps the most fundamental error in the multiplexing process would be an error in the phase of the main timing reference, burst. Since the entire system is based on burst phase, an error in burst phase will appear as an opposite error in every phase except burst, because the circuits will insist that burst phase cannot be wrong. The general result will be an overall hue error in the reproduced picture. This effect can be better visualized by referring to Figure 1.9-43.

A phase error in burst produces the same result as holding burst phase stationary and allowing all other phases to slip around the circle an equal amount (but in a direction opposite to the burst phase error). Each color vector then represents a hue other than the one intended.

Burst Amplitude Error

In theory, the receiver circuits which extract timing information from the burst are insensitive to variations in burst amplitude as long as the burst is large enough to maintain a respectable signal-to-noise ratio and not so large that some type of clipping or rectification upsets the burst circuitry. But practical receivers always exhibit some degree of sensitivity depending mainly upon the error in the subcarrier oscillator in the receiver. If the free-running frequency of the receiver oscillator is very different from burst frequency—particularly if the difference is so great that the burst is in danger of losing control of the oscillator—then a fairly appreciable amplitude sensitivity will be noted. This sensitivity will take the form of a phase error, and the net result will be indistinguishable from a burst phase error, as previously discussed.

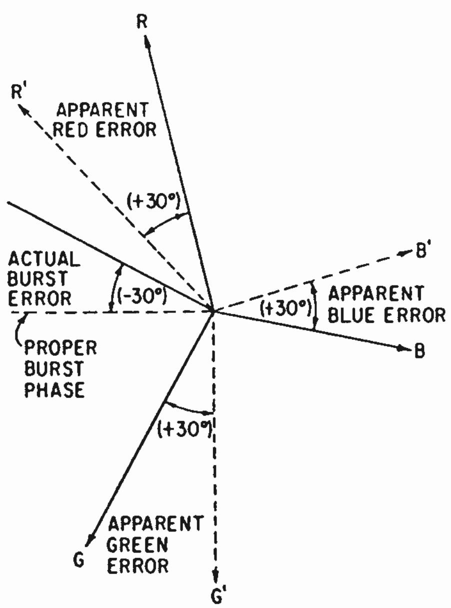


Figure 1.9-43. Vector diagram showing how error in subcarrier phase becomes an opposite error in all other phases.

Some receivers have a circuit which automatically adjusts the gain of the color information channels so that the viewer always sees the proper saturations, regardless of errors, which might tend either to wash out or oversaturate the picture. Such a circuit, called an *automatic chroma control* (ACC), derives its control information from the amplitude of burst, which is presumed to bear a constant ratio to the amplitude of chroma. Transmission distortions, for example, might decrease the amplitude of both burst and chroma, but since the ratios of their amplitudes would be preserved, an ACC receiver could automatically modify its chromachannel gain to compensate for the decreased chroma amplitude. However, if a color encoder error should cause burst alone to decrease in amplitude, the ACC circuits would increase chroma gain just as in the previous case, with the result that a viewer would receive an oversaturated picture.

Two-Phase Modulation Errors

The fidelity of color reproduction can be seriously affected if the phase separation of the Q and I subcarriers is not maintained at 90°. It can be shown that a slip in the angular position of the Q axis, for example, will result in cross talk of Q and I. The final result will be the same as cross talk among the primary colors.

Likewise, in a receiver, the phase relationship between the reference subcarriers must be maintained to avoid a similar error. Any deviation from the proper phase relationship will have a similar result, that is, cross talk of I into Q or Q into I, with the net picture result resembling cross talk among all the primary colors.

Carrier Unbalance

In a properly operating doubly balanced modulator, the carrier component of the signal is suppressed in the modulator circuit. If some error in components or operation causes this suppression to be imperfect, the carrier will appear in the output. This condition is known as *carrier unbalance*.

The effect of carrier unbalance can be evaluated by considering the unwanted carrier as a vector of constant amplitude which adds itself vectorially to every vector present in the colorplexer output. In general, such a vector will shift all vectors and hence all hues seen in the picture toward one end of the other of the color axis represented by the unbalanced modulator. For example, a positive unbalance in the I modulator would shift all colors toward the color represented by the positive I axis, that is, toward orange. A negative I unbalance would shift all colors toward cyan.

To visualize this effect, refer to Figure 1.9-44, in which has been added to each color vector a small positive vector which is parallel to the I axis. This small vector represents the amount of carrier unbalance. The resultant vectors will all be rotated toward the positive I axis and changed in amplitude as well. Such changes represent errors in both hue and saturation.

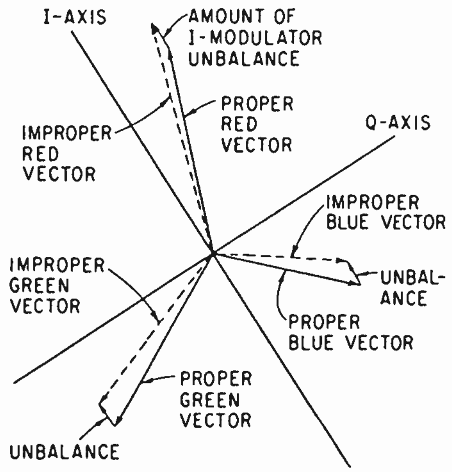


Figure 1.9-44. Vector diagram of subcarrier phase and amplitude with positive vectors added to represent carrier unbalance in the I modulator.

Another error from carrier unbalance occurs in white and gray areas of the picture. In a normally operating colorplexer, a white (or gray) area in the scene causes the Q and I signals to become zero and thereby causes the modulator outputs to become zero. Hence, a white or gray area will normally appear in the signal as an interval of zero subcarrier amplitude. If one of the modulators begins to produce a carrier unbalance vector, however, a white or gray area will become colored because of the subcarrier which will be added in this interval. Moreover, certain areas, which are normally colored, may have their subcarrier canceled by the carrier unbalance vector and become white. Such white-to-color and color-to-white errors are very objectionable.

Video Unbalance

A doubly balanced modulator derives its name from the fact that it balances out or suppresses both the carrier and the modulating video (Q or I). If, for any reason, the video suppression becomes less than perfect, the resulting condition is called *video unbalance*.

Video unbalance will cause unwanted Q or I video to appear in the modulator output, in addition to the desired sideband outputs. This unwanted video signal will be added to the luminance signal, thereby distorting the gray scale of the picture. For example, a slight positive unbalance in the Q modulator would slightly brighten reds and blues and slightly darken greens. A negative unbalance would have the opposite effect.

Subcarrier Frequency Error

The color subcarrier frequency is specified by the FCC to be 3.579545 MHz ± 10 Hz. Deviations within this specified limit are of no consequence (provided they are slow deviations). Large deviations, however, can affect color fidelity. The effect does not usually become serious within the possible frequency range of a good crystal-controlled subcarrier source driving a properly designed receiver.

In receivers, the subcarrier timing information is extracted from the burst on the back porch and used to control the frequency of a subcarrier-frequency oscillator in the receiver. As long as the unlocked frequencies of the burst and the receiver oscillator remain the same, the locked phase relationship between the two will remain the same. But if either the burst frequency or the receiver-oscillator frequency becomes different (and the difference between them is not so large that lockup is impossible), then the locked error, which obviously cannot be a frequency error, manifests itself as a phase error. This error can become as large as ±90° before the AFC circuit can no longer hold the receiver oscillator on frequency. The frequency range over which this phase shift occurs depends upon the receiver design.

Possible Distortions in the Transmission System

Preceding sections have described the processes involved in the generation and display of a color television signal. Errors in these processes are not the only possible source of distortion; when the signal is transmitted over great distances, the transmission system itself may contribute errors. This section discusses parameters which specify the behavior of a transmission system and describes the effects that errors in these parameters can have on the reproduced picture.

This section is divided into two parts. The first relates to the parameters of a perfectly linear transmission system, while the second part discusses the additional parameters required to describe the nonlinearities that are inevitable in any practical system.

The Perfectly Linear Transmission System

A perfectly linear and noise free transmission system can be described by its gain and phase characteristics plotted against frequency as the independent variable.[2] Typical plots are shown in Figures 1.9-45 and 46, respectively. These two characteristics known, it is possible to predict accurately what effect the transmission system will have on a given signal.

Gain Characteristic

Figure 1.9-45 is usually known as the frequency response or gain characteristic of the system. Ideally,

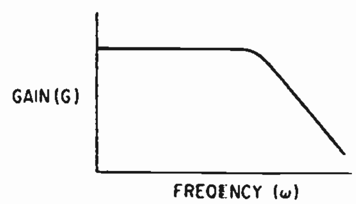


Figure 1.9-45. Typical curve showing a gain of a system plotted against frequency to determine its gain characteristic.

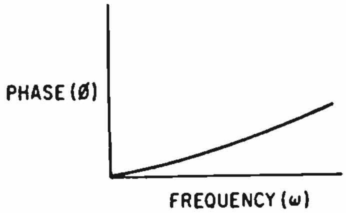


Figure 1.9-46. Curve showing phase characteristic of a system plotted versus frequency.

it should be perfectly flat from zero to infinite frequency, but this, of course, is impossible to attain. An amplifier has a definite gain bandwidth product, depending upon the transconductance of its active elements (transistors, ICs, etc), the distributed capacity shunting these elements, and the types of compensation (peaking) employed. The bandwidth of a given combination of transistors, stray capacitances, and peaking networks can be increased only by decreasing its gain, or conversely, its gain can be increased only by decreasing its bandwidth. There is a limitation, therefore, to the actual bandwidth that can be obtained. For a given scanning standard, the bandwidth required in a monochrome television system is determined by the desired ratio between the horizontal resolution and the vertical resolution. Although nominally a 4.0 MHz bandwidth is required for the monochrome standards, the requirement can be relaxed to the detriment of only the horizontal resolution. The subjective result is a softening of the picture in proportion to the narrowing of the bandwidth (neglecting the influence of the phase characteristic in the vicinity of the cutoff frequency). As pointed out in preceding sections, the entire chrominance information of the color system is located in the upper 1.5 MHz of the prescribed 4.0 MHz channel; hence, any loss of response in this part of the spectrum can have a marked effect on the color fidelity of the reproduced picture.

One of the most serious forms of distortion inflicted on a color picture by bandwidth limiting is loss of saturation. Consider a case in which the bandwidth is so narrow as to result in no gain at the color subcarrier frequency. The output signal then contains no color subcarrier and hence reaches the color receiver as a monochrome signal, producing zero saturation. Nearly as poor results can be expected from an amplifier with response such that the gain at 3.58 MHz is one-half the low-frequency gain. Since the saturation depends chiefly on the amplitude of the subcarrier, the saturation will be correspondingly reduced. The resultant color picture will have a washed out look.

Loss of high-frequency response, which can be expected to contribute to loss of fidelity, is usually accompanied by phase disturbance, depending on the type of networks employed in the system. The intent in this section, however, is to treat each variable separately. Therefore, discussions are based on the effects of varying only one parameter of a system. It is suggested that the reader can determine the combined effect of two or more variables by comparing the results shown for the individual variables.

Phase Characteristic

An ideal system has a *linear* phase characteristic, as shown in Figure 1.9-47(a). Such a characteristic implies that all frequencies of a signal have exactly the same *time delay* in passing through this system, since the time delay is given by the phase angle divided by the (radian) frequency. It can be seen in Figure 1.9-47 that if three frequencies are chose arbitrarily, then the corresponding phase angles must have values proportional to their corresponding frequencies (because of the geometric properties of a right triangle). To state it another way, if $\varphi_1/\omega_1 = 0.2$ μsec, then $\varphi_2/\omega_2 = 0.2$ μsec and φ_3/ω_3 also equals 0.2 μsec. Plotting these three values and drawing a straight line through them as in Figure 1.9-47(b) will show that the time delay for all frequencies is 0.2 μsec.

A signal is not distorted by delay as long as all parts of it are delayed by the same amount. However, when the phase characteristic is nonlinear (as in Figure 1.9-48(a)), the time delays for all parts of the signal are no longer equal (see Figure 1.9-48(b)). For example, if a complex waveform is made up of a 1 MHz sine wave and its third harmonic, these two components will suffer unequal delays in passing through a system having the characteristics of Figure 1.9-47. The resultant distortion can be seen by comparing Figures 49(a), 49(b) and 49(c).

Such distortion is detrimental to both the luminance and chrominance of a composite signal. The luminance signal will have its edges and other important details *scattered,* or *dispersed,* in the final image. Such a transmission system is said to introduce *dispersion.* (Conversely, if a system does not scatter the edges and other high-frequency information, it is said to be

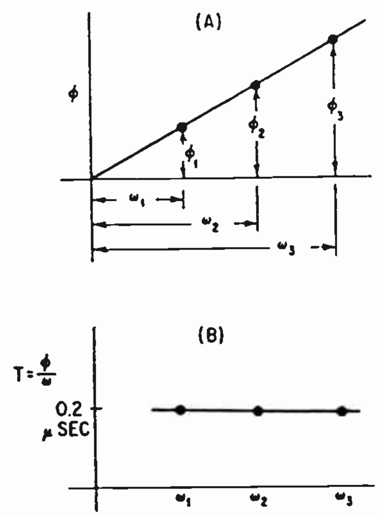


Figure 1.9-47. Curves illustrating a system with linear phase characteristics, which will give the same time delay for signals of all frequencies.

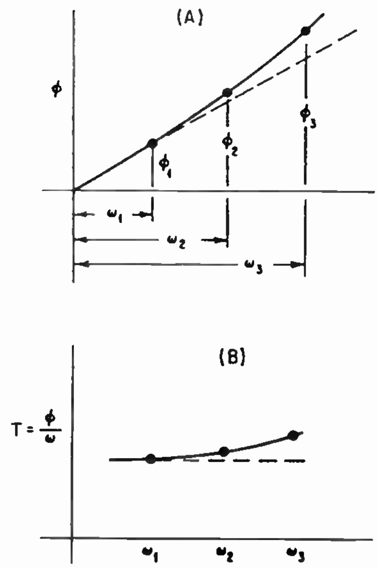


Figure 1.9-48. Curves showing the effect of nonlinear phase characteristic on time delay characteristic.

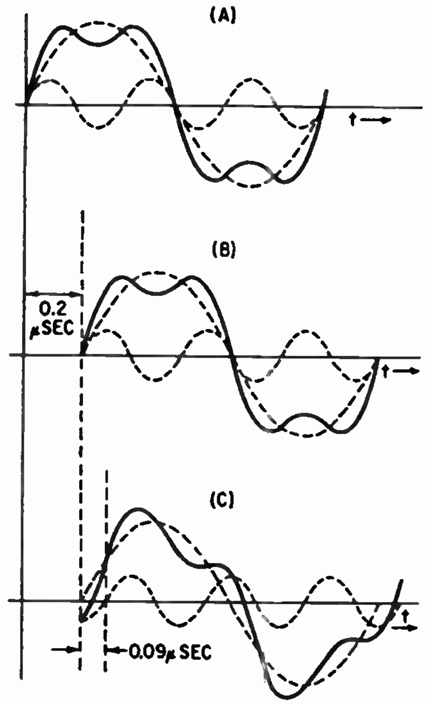


Figure 1.9-49. Curves showing that a complex wave (a) is not distorted by time delay (b) when both components (shown dotted) are delayed by the same amount. Unequal delays (c), however, cause distortion.

dispersionless.) The effect of phase distortion on the chrominance information is of a rather special nature and can best be explained by introducing the concept of *envelope delay.*

Envelope Delay

In the preceding discussion, the time delays φ_1/ω_1, φ_2/ω_2 and φ_3/ω_3 were always determined by measuring the frequencies and the phases from $\varphi = 0$ and $\omega = 0$. It might be said that the delay at zero frequency is commonly taken as the reference point for all other delays. This method is usually adequate for determining the performance of systems that do not carry any signals which have been modulated onto a carrier. But a carrier, with its family of associated sidebands (Figure 1.9-50(b)), can be thought of as a method of transmitting signals in which the zero frequency reference is translated to a carrier frequency reference. This translation can be understood by referring to Figures 1.9-50(a) and 50(b). To calculate the delay of the carrier borne signals *after* they have been demodulated, measurements of φ and ω must be referenced, not from zero frequency, but from *carrier* frequency.

In Figure 1.9-51(a), an impossible phase characteristic has been drawn to aid in further discussion of this subject. Such a characteristic, consisting of two perfectly straight lines, is never met in practice but makes a very simple system for developing the subject of envelope delay.

First, pass two frequencies ω_1 and ω_2 through this system. Let ω_1 be a carrier and ω_2 a sideband which might be, for example, 1,000 Hz higher. If ω_1 and ω_2 fall on the characteristic as shown in Figure 1.9-51(a), the delay which the 1,000 Hz will show after demodulation can be found by putting new reference axes (shown dotted) with ω_1, the carrier, at zero on these new axes. Now, when ω_s and φ_s are measured as shown, the time delay after demodulation is φ_s/ω_s. In this case, the delay of the 1,000 Hz after demodulation is the same as it would have been had it been passed through the system directly.

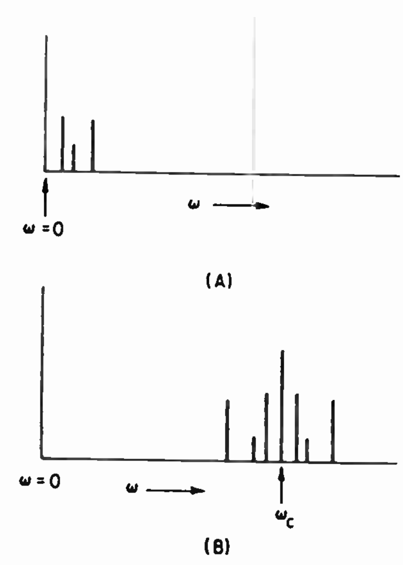


Figure 1.9-50. Sketch showing how a group of frequencies near $\omega = 0$ [sec. (a)] can be translated by modulation onto a carrier to a group of sidebands near ω_c – a carrier frequency (b).

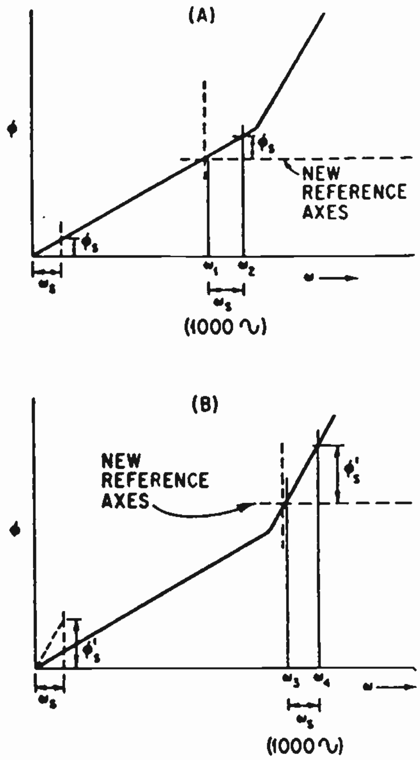


Figure 1.9-51. Idealized straight-line phase characteristics showing how a carrierborne 1,000 Hz signal can be delayed excessively when the carrier and sideband fall on a steeper portion of the phase characteristic.

Second, pass two other frequencies ω_3 and ω_4 through this system as redrawn in Figure 1.9-51(b). This time drawing in the new axes at ν_3, it can be seen that although ω_s is still 1,000 Hz, φ_s is larger than φ_s. Therefore, it can be concluded that the time delay φ_s/ω_s for this second case is greater than for the first case. The 1,000 Hz, when demodulated, will show a considerable error in timing.

Stressing the phrase *delay in a demodulated wave* should not be taken to mean that the demodulation process produces this delay or even make it apparent where it was previously not detectable. Any delay that a demodulated wave shows was also present when the wave existed as a carrier having an envelope. In short, the delay of the demodulated wave appears first as a delay of the envelope, hence the term *envelope delay.*

Envelope delay does not constitute a distortion. If a system such as the one shown in Figure 1.9-51(a) introduces a delay of 0.2 μsec to the 1,000 Hz wave (measured after demodulation), then the envelope delay of the system is 0.2 μsec. However, it was shown that a 1,000 Hz signal passed directly through the system (without first being modulated into a carrier) would also suffer a delay of 0.2 μsec. As long as the envelope delay φ_s/ω_s is the same as the time delay φ_1/ω_1 the envelope delay introduces no timing errors. But in the second system (Figure 1.9-51(b)) the demodulated 1,000 Hz wave suffered a *larger* delay, say 0.29 μsec. A 1,000 Hz signal passed directly through this system, however, would still be delayed only 0.2 μsec. Therefore, the second system has an envelope delay of 0.29 μsec and an *envelope delay distortion* of 0.09 μsec.

It is probably wise to point out that the time delay φ_3/ω_3 in Figure 1.9-51(b) is considerably less than the 0.29 μsec estimated for the value of envelope delay. Although φ_3/ω_3 would be greater than 0.2 μsec (say, for example, that φ_3/ω_3 is 0.22 μsec), the value would be optimistic about the amount of timing error that would be shown by the demodulated 1,000 Hz signal. The need for knowledge of the envelope delay φ_s/ω_s of the system is therefore obvious.

Effect of Envelope Delay Distortion on a Color Picture

A transmission system which exhibits envelope delay distortion will destroy the time coincidence between the chrominance and luminance portions of the signal. This will result in misregistration between the color and luminance components of the reproduced picture. The following paragraph explains briefly how envelope delay distortion causes this error.

Any colored area in a reproduced picture is derived from two signals: a chrominance signal and a luminance signal. Since these two signals describe the same area in the scene, they begin and end at the same time. The chrominance signal arrives at the receiver as a modulated subcarrier; the luminance signal does not. Therefore, as previously shown, the delay of the chrominance signal is determined principally by the envelope delay of the system and the delay of the luminance signal is determined principally by the ordinary time delay φ/ω. If the two delays are not identical (that is, if there is envelope delay *distortion*), then the chrominance signal does not coincide with the luminance signal and the resultant picture suffers *color luminance misregistration* in a horizontal direction.

For example, in a system having the characteristic of Figure 1.9-51(b), the luminance signal is delayed by 0.2 μsec but the chrominance signal is delayed by 0.29 μsec. The error in registration then amounts to 0.09 μsec, or about 0.2% of the horizontal dimension of the picture, which is about 0.3 in. on a 21-in. (diagonal) picture.

Although the subject of compatibility is outside the scope of this chapter, it is worth noting in passing that envelope delay distortion adversely affects compatibility, since it causes wideband monochrome receivers to display a misregistered dot-crawl image in addition to the proper luminance image.

General Method for Envelope Delay

The specific cases described (Figures 1.9-51(a) and 51(b)) made use of simple, idealized straight line approximations to develop the concept of envelope delay. Practical circuits are not so simple. For example, a simple *RC* network has a φ versus ω plot as in Figure 1.9-52. Finding the envelope delay of this curved line plot will clarify what is meant by envelope delay.

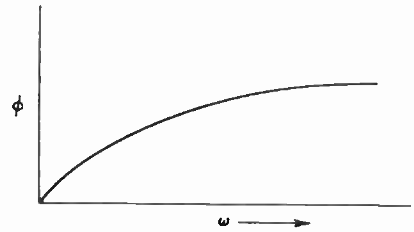


Figure 1.9-52. Phase characteristic of an RC network.

Referring back to the plots of Figure 1.9-50(a) and 50(b), it can be seen that the characteristic of the plot that determines the value of envelope delay is its *slope*. The larger envelope delay, which was suffered by the ω_3-ω_4 pair (Figure 1.9-51(b)), was a result of their lying on the steeper slope. The envelope delay of *any* system is equal to the slope of the phase versus frequency characteristic. If this characteristic is a curved line (as for the *RC* network, Fig. 1.9-52), then the slope is different at every frequency.

The slope of a curved line can be found by the methods of the differential calculus or to a good approximation by breaking up the line into a number of straight-line segments, as in Figure 1.9-53. If the slope of each of these straight lines is plotted against its corresponding frequency (that corresponding to the center of the line), the resulting curve will be approximately the envelope delay characteristic.

Nonlinearities of a Practical Transmission System

It is important to emphasize that the effect of nonlinearities in a color television system depends upon whether these nonlinearities precede or follow the matrixing and modulation sections of the system. Nonlinearities in transfer characteristics detract from color fidelity; the same degree of nonlinearity after matrixing and modulation also affects color fidelity although in a different way. The purpose of the following paragraphs is to discuss how a nonlinear transmission system affects a *composite* color signal. It is assumed that all other nonlinearities in the entire system either are negligible or have been canceled by use of nonlinear amplifiers such as gamma correctors.

The major sources of nonlinearity in a transmission system are its amplifying devices. These devices have a limited dynamic range. For example, if too much signal is supplied to them, an *overload* results. The transfer characteristic of such a system can be sketched as in Figure 1.9-54(a).

Such a nonlinearity is one of three types commonly encountered in video transmission systems. These three types are:

- Incremental gain distortion
- Differential gain
- Differential phase

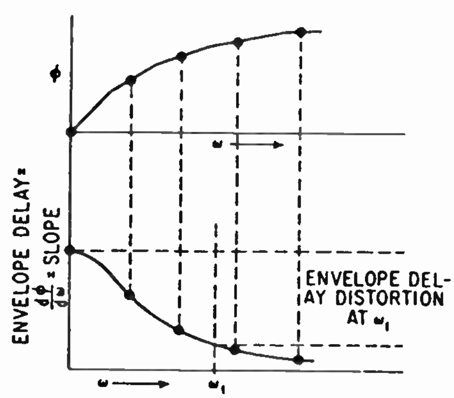


Figure 1.9-53. Graphs showing how a series of straight-line segments can be used to approximate the envelope delay characteristics (bottom).

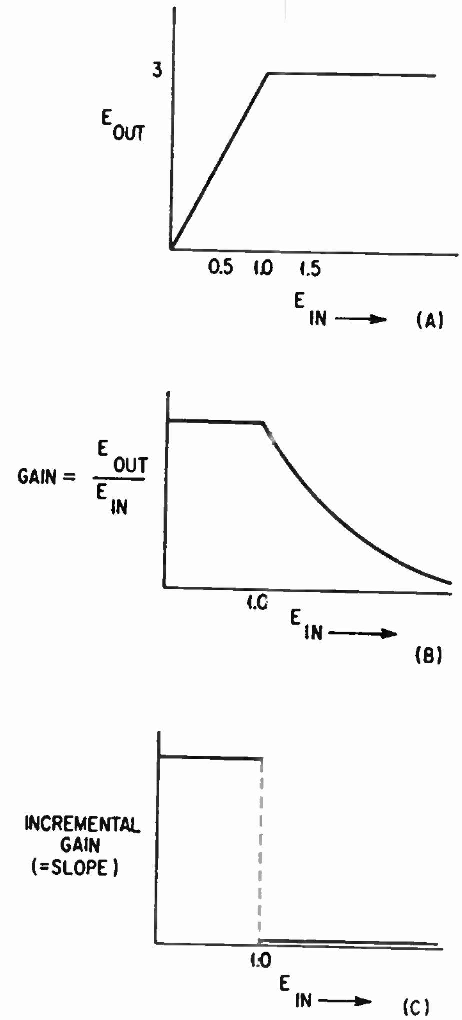


Figure 1.9-54. Idealized straight-line plots showing (a) output voltage of an amplifier versus input voltage (b) gain of the amplifier versus input voltage and (c) incremental gain of the amplifier versus input voltage. Curve (c) is the slope of curve (a).

Incremental Gain

The concept of the slope of a plot, developed in the discussion of envelope delay, will be useful here as well. Consider a plot as in Figure 1.9-54(a) which shows output voltage of an amplifier plotted against input voltage. Idealized straight-line plots are shown for simplicity. It can be seen that the amplifier has a maximum output of 3 V for 1 V input. Larger input voltages result in no more output; the amplifier *clips* or *compresses* when inputs larger than 1 V are applied.

The gain of the amplifier is

$$\text{Gain} = \frac{E_0}{E_{in}} = \frac{3\text{ volts}}{1\text{ volt}} = 3$$

The gain is obviously constant below the clip point. For example, *k* an input voltage of 0.5 V gives

$$\text{Gain} = \frac{1.5\text{ volts}}{0.5\text{ volts}} = 3$$

But at an input of 1.5 V, the output is still 3 V, so the gain is only 2. (The word "gain" is of doubtful use here because of the clipping involved.) The gain, defined as E_0/E_{in}, is plotted against E_{in} in Figure 1.9-54(b). It can be seen in this figure that the gain is constant only as long as the *slope* of Figure 1.9-54(a) is constant.

It is useful, then, to establish a new term, *incremental gain*, which will be defined as the *slope* of a plot such as Figure 1.9-54(a). For the particular plot of Figure 1.9-54(a), the slope is constant up to E_{in} = 1 V and then suddenly becomes zero. The corresponding plot of slope versus E_{in} is shown in Figure 1.9-54(c).

The importance of incremental gain in color television can be assessed by applying the input signal shown in Figure 1.9-55 to the distorting system of Figure 1.9-54(a). Before being applied to the distorting system, such a signal could be reproduced on a monochrome receiver as a vertical white bar and on a color receiver as a pastel colored bar, say, for example, a pale green. After passing through the distorting system, the signal would still be reproduced as a white bar on the monochrome receiver with the only apparent error being a luminance distortion, that is, a slight reduction in brightness, which, for the magnitudes shown here, would probably pass unnoticed. The color receiver, however, would receive a signal completely devoid of any color information and would reproduce a white bar in place of the former pale green one.

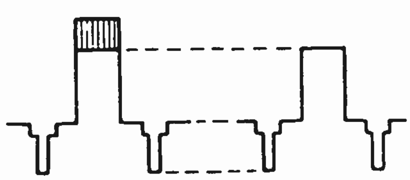


Figure 1.9-55. Extreme case of distortion resulting from passing signal at left (a) through the amplifier represented by Figure 1.9-54. The output (b) has no color information remaining.

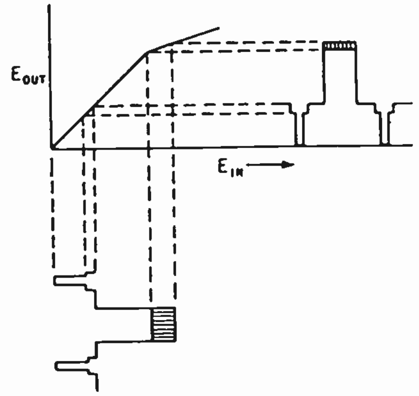


Figure 1.9-56. Diagram showing effect of incremental gain distortion of reducing amplitude of color portion of signal.

The less extreme case is shown in Figure 1.9-56. For the system represented by this characteristic, the slope (incremental gain) does not become zero for inputs above 1 V but instead falls to one-half its below-1 V value. The color signal of Figure 1.9-56 would not lose all color in passing through this system, but the amplitude of the subcarrier would become only one-half of its proper value. Since saturation is a function of subcarrier amplitude, the pale green of the undistorted reproduction would, in this case, become a paler green. The luminance distortion would also be less than in the extreme (clipping) case.

It can be seen, then, that unless the incremental gain of a system is constant, that system will introduce compression which will distort the saturation and brightness of reproduced colors. Usually, the error is in the direction of *decreased* luminance and saturation. For certain systems, however, exceptions can be found. For example, the effect that the system represented by Figure 1.9-56 will have on a signal depends on the polarity of the signal. For the signal as shown, the usual decrease in luminance and saturation is exhibited. For an inverted signal, however, the subcarrier amplitude would not be reduced, but the luminance signal would still be diminished. The subjective result of this distortion would be an *increase* in saturation. The unusual behavior of this particular system is attributable to its peculiar transfer characteristic, which was drawn with curvature at one end only to simplify the discussion. Most practical system transfer characteristics exhibit curvature at both ends and therefore have an effect on the signal, which is essentially independent of polarity.

Incremental gain can be measured in two ways, the first of which stems from its contribution to luminance distortion and the second, from its contribution to chrominance distortion.

In the first method, an equal-step staircase waveform such as shown in Figure 1.9-57(a) is applied to the system to simulate a signal having equal luminance increments. If the system has constant incremental gain, the output will, of course, also have equal-step increments. But if the system does not have constant

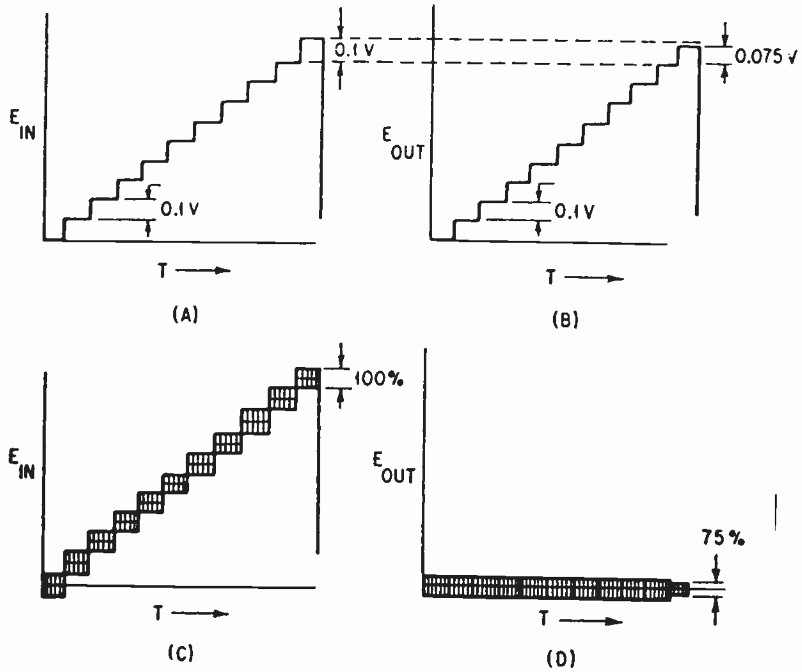


Figure 1.9-57. Diagrams showing two methods of measuring incremental gain distortion, namely, in (a) and (b) by its contribution to luminance distortion and in (c) and (d) chrominance distortion.

incremental gain, certain of the steps will be compressed, as in Figure 1.9-57(b). If the compression is as in the Figure, the *incremental gain distortion* (IGD) is indicated by the distorted amplitude of the last step. Numerically, it can be stated as a percentage:

$$\text{IGD} = 1 - \frac{S_{distorted}}{S_{undistorted}} \times 100\%$$

where S is a step amplitude.

For example, if an undistorted step is 0.1 V and the distorted one is 0.075 V, then the incremental gain distortion would be 25%.

Using the chrominance distortion technique, an input signal consisting of the step wave plus a small, high-frequency sine wave, as shown in Figure 1.9-57(c), is applied to the system. After the signal has passed through the system, it is fed through a high pass filter, which removes the low-frequency staircase. The incremental gain distortion then is indicated by the differences in the amplitude of the high-frequency sine waves (see Figure 1.9-57(d)). In this case, the high-frequency sine wave associated with the top step is shown as having 75% of the amplitude of the sine waves associated with the lower steps, which are assumed to be undistorted. Again, the incremental gain distortion is 25%.

A most important point must be made regarding the equivalence of these two techniques. Certain systems which show incremental gain distortion when tested by the luminance-step technique may or may not show the same distortion when tested by the high-frequency and high pass filter technique. Moreover, a system which shows distortion by the second technique may or may not show distortion by the first. In other words, the incremental gain distortion may be different for different frequencies. Such differences are frequently found in staggered amplifiers, feedback amplifiers or amplifiers having separate parallel paths for high and low frequencies, such as might be found in stabilizing amplifiers.

A thorough test of a system, therefore, should include a test of its incremental gain by both techniques. The staircase-plus-high-frequency waveform can be used to provide *both* tests by observing the system output (for this test waveform input) first through a low pass filter and then through a high pass filter. The first test will show low-frequency distortions; the second, high-frequency distortions.

Differential Gain

On the basis of the previous discussion of incremental gain distortion, the extremely important concept of *differential gain* can be presented merely as a simple definition. Differential gain is identical with incremental gain distortion when the latter is measured by observing ". . . the difference in the gain of the system for a small high-frequency sine-wave signal at two stated levels of a low-frequency signal upon which it is superimposed."[4] In other words, differential gain is a special form of incremental gain distortion, which describes the IGD of a system for the superimposed high-frequency case only.

One of the reasons for selecting the high-frequency aspect of incremental gain distortion for the Institute of Radio Engineers (IRE) definition of differential gain was applied in Figure 1.9-55, when the "... high-frequency sine wave ..." of the definition was made equal to color subcarrier. This special case of differential gain explores the system gain linearity in the vicinity of this particularly important frequency. The definition of differential gain was purposely made in the broad terms of a "... high frequency sine wave ..." to allow the greatest possible versatility in devising methods of measurement. In present color television practice, however, the "... high frequency sine wave ..." is always color subcarrier and the low-frequency signal mentioned in the definition is a 15,750 Hz staircase, sine wave, or sawtooth.

Another reason for emphasizing high-frequency IGD was implied previously by the sentence "... the signal ... would ... be reproduced ... with the only apparent error being a luminance distortion ... which, for the magnitudes shown here, would probably pass unnoticed." The magnitude shown was a 25% IGD, which is passing unnoticed, indicating that large incremental gain distortions usually cause no detectable luminance errors. Incremental gain distortion is almost too sensitive a tool to measure luminance distortions. For this purpose, simple gain distortion (compression) is more useful. Therefore, the luminance/distortion aspect of IGD was deliberately omitted from the definition of differential gain.

Incremental Phase and Differential Phase

The phase characteristic sketched in Figure 1.9-46 indicates that the system described by this plot will introduce a certain amount of phase shift for any given frequency. For example, it might be found that a certain system would introduce a phase shift of 60° at 2 MHz. If the system in question were perfectly linear, this 60° phase shift would be produced regardless of how the 2 MHz signal might be applied to the system.

It can be shown, however, that some systems, when presented with a signal of the type shown in Figure 1.9-58, will introduce a delay different from 60°, depending on where the zero axis of the sine wave falls on the transfer characteristic of the system. For the case sketched in the figure a phase shift of 70° is drawn for the largest zero-axis displacement.

By analogy with the incremental gain and differential gain arguments, it is possible to define three quantities which pertain to this type of distortion. These quantities are *incremental phase, incremental phase distortion,* and *differential phase.* It can also be shown that of the three, differential phase is the most important quantity.

Incremental phase is the least exact analogue, since it is not very similar in form to incremental gain. Incremental *gain* is a *slope*; incremental phase is simply the absolute value of phase shift. In the previous system, the incremental phase was 60° or 70° (or somewhere in between), depending upon the location of the zero axis.

Incremental phase distortion, like its analogue (*incremental gain distortion*), depends upon the magnitude of the error. It should be zero for a perfect system. In the system of Figure 1.9-58 the 2 MHz signal with 70° incremental phase would be said to have 10° incremental phase distortion, so it is clear that the difference between two phases (one of which is assumed to be correct) gives the incremental phase distortion.

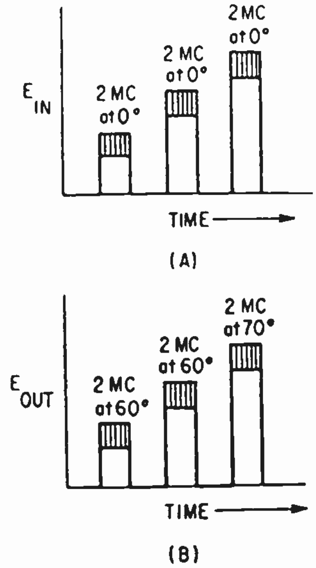


Figure 1.9-58. Graphs illustrating how a signal (a) may undergo different phase shifts (b) depending upon where the zero axis at the sine wave falls on the system transfer characteristic. This distortion is called differential phase.

As previously stated, *differential gain* is identical with incremental gain distortion for the superimposed high-frequency case only. Similarly, differential phase is identical with incremental phase distortion, but there is no need to limit the definition to the superimposed high-frequency case, since there is no other case which is meaningful for phase distortion. Without the superimposed sine wave, no phase measurement is possible. Therefore, differential phase is identical with incremental phase distortion. In practical work, the first two terms are seldom used, for the last, differential phase, has been found completely adequate to describe this aspect of a system.

In summary, the differential phase of a system is "the difference in phase shift through the system for a small high-frequency sine wave signal at two stated levels of a low-frequency signal on which it is superimposed."[5]

Effect of Differential Phase on Color Picture

The phase of a subcarrier in a composite signal carries information about the *hue* of the signal at that instant. If the signal passes through a system which introduces differential phase, the subcarrier phase (and hence, the hue) at the output will become dependent upon the amplitude of the luminance associated with the hue, since it is the luminance signal which determines the location of the zero axis of the subcarrier.

For example, a system introducing 10° of differential phase might be adjusted to reproduce properly a low-luminance hue such as saturated blue or a high-luminance hue such as saturated yellow, but *not both*. One or the other would have to be in error.

Conclusion

This discussion of color errors indicates possible degradations in color fidelity and their probable sources. However, in a properly adjusted color TV system the picture quality is excellent. The various techniques now in use within the framework of the NTSC system have assured a long future for color TV, even in the DTV era now upon us.

THE COLOR ENCODER

The color encoder in the color television system performs the required encoding of the R, G, B signals from three-tube cameras or the R, G, B and Y (luminance) signals from four-tube cameras into a single color video signal conforming to FCC specifications. It is the heart of the modern color television system and represents a most ingenious application of many elements of communication circuit theory. Figure 1.9-59 shows a block schematic of a basic color television system indicating the functions and major components of the color encoder.

A more detailed block diagram of the color encoder showing the matrixing, bandwidth limiting and quadrature modulation functions is shown in Figures 1.9-33 and 34.

Basic Functions

The principal operations and functions performed by the color encoder are:

- Matrixing of R, G, B video signals to produce luminance and chrominance signals
- Filtering of the chrominance signals to obtain the required bandwidth
- Delay compensation to correct for band-limiting time delay
- Modulation of 3.58 MHz carriers by chrominance signals
- Insertion of color sync burst
- Addition of luminance and chrominance signal to form a complete color signal
- Optional addition of sync.

Color encoders of modern design are inherently stable and require only routine verification or adjustment. Set-up of a color encoder involves the use of color bars which are electrically generated waveforms of high precision. A color bar generator is capable of producing on a color monitor all of the signal bars illustrated in Figure 1.9-60.

Colors at the top of this display pattern are arranged from left to right as white, yellow, cyan, green, magenta, red and blue in their decreasing order of luminance. The lower portion of the pattern contains I, 100% white, Q, and black signal areas. The I and Q signals simplify subcarrier phase adjustments in the color encoder and the 100% white bar facilities white balance adjustments. The specifications of the standard encoder color bar signal are given in EIA Standard RS-189.

Waveforms

Figure 1.9-61 shows the oscilloscope waveforms at a horizontal sweep rate of the color bar signals displayed on the television raster. Note that this is a composite representation of waveforms of the top and

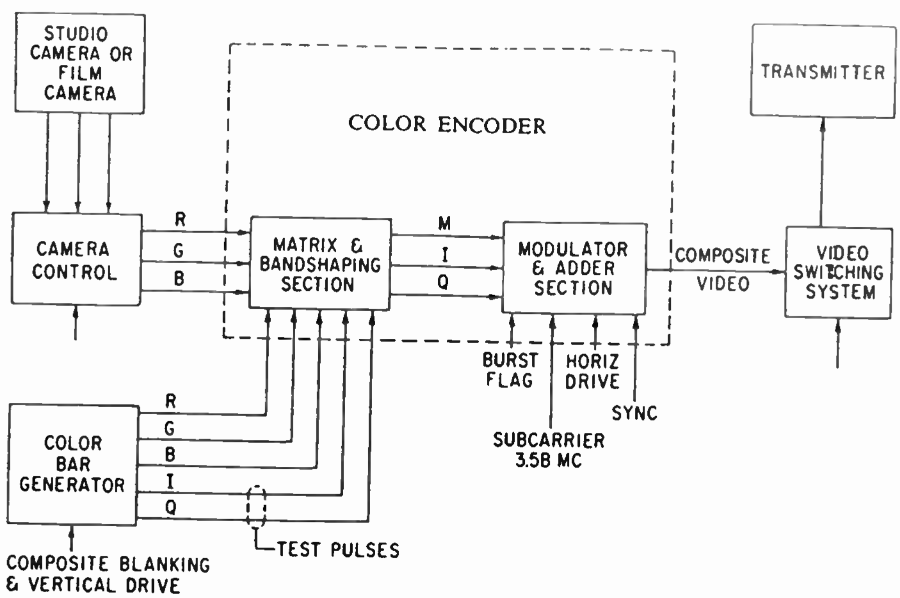


Figure 1.9-59. Basic color television system showing functions and major components of the color encoder.

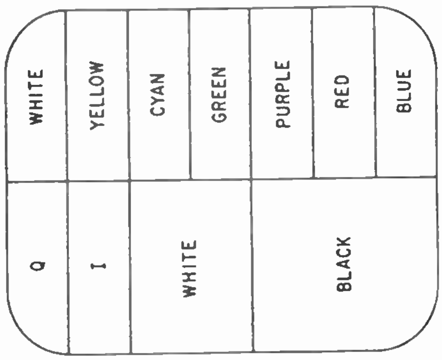


Figure 1.9-60. Diagram showing color monitor display of color and test bars electronically produced by RCA color-based generator.

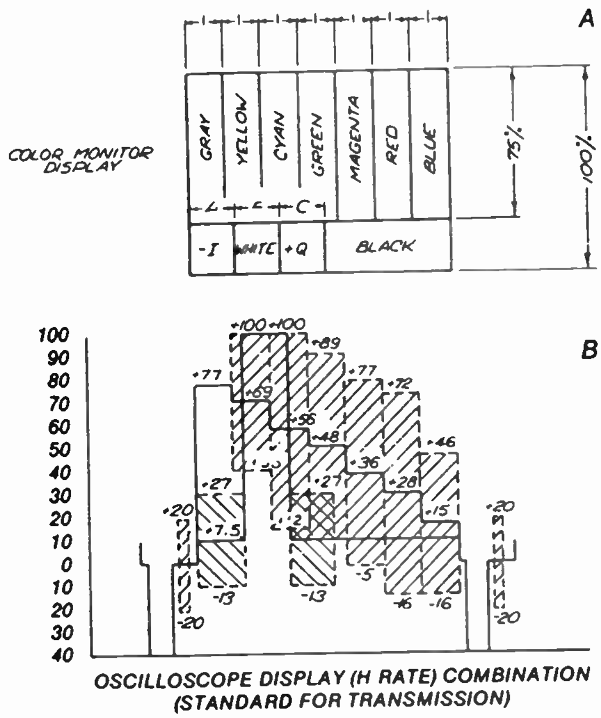


Figure 1.9-61. (a) Color monitor display and (b) Oscilloscope display (H rate).

bottom areas of the raster. The color sync precedes the color bar pulse information.

Figure 1.9-62 shows the various band-pass response characteristics of the luminance channel and of the I and Q channels of the color encoder.

A color encoder is set up and adjusted by using the calibrated color bars just described. The color encoder luminance gain is adjusted by using the 75% white bar as a reference. By switching off the luminance channel the appropriate I and Q waveforms are available to set the proper peak amplitudes and the 90° phase separation. Either a wide-band oscilloscope or a vectorscope can be used for display in a variety of specialized setup procedures. The vector relationship of chrominance components is shown in Figure 1.9-63.

COLOR TEST EQUIPMENT

The color television broadcast station relies heavily on specialized test and monitoring facilities in order to maintain adequate standards of performance and to ensure compliance with FCC regulations. Test signals have become quite sophisticated and yield much useful information on the performance of the systems.

A stable high performance color monitor is an essential element of color test equipment. This, together with a vectorscope and a standard color bar generator for setup and calibration serves as a means of evaluating performance.

The color monitor, vectorscope and color bar generator find utilization in a rapid routine day-to-day check of the television system adjustments.

Additional test equipment needed for color TV performance evaluation falls into two categories: equipment to evaluate studio performance and equipment to evaluate microwave relay and transmitter performance.

The important electrical characteristics to be measured in either category are:

- Linearity or differential gain
- Frequency response and differential phase performance
- Group delay characteristic
- Low frequency square-wave response.

Evolutionary developments have followed the requirement that specific test waveforms be made available which are compatible with normal television sig-

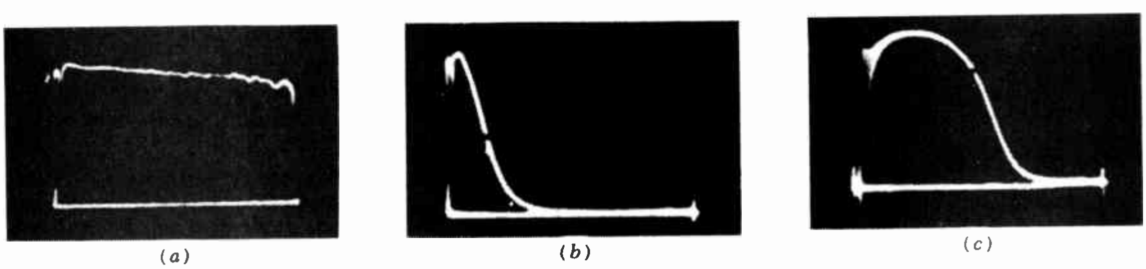


Figure 1.9-62. Waveforms showing response characteristics of monochrome, I and Q channels. (a) Response of monochrome channel without aperture correction, marker at 8.0 MHz (b) output if I filter, marker at 2.0 MHz; (c) output of Q filter, marker at 500 kHz.

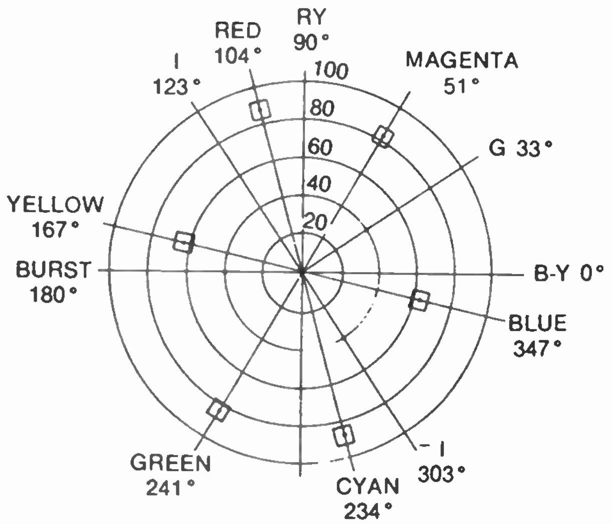


Figure 1.9-63. Vector relationship among chrominance components.

nal systems and can be introduced easily without disabling or upsetting normal operating conditions. Measurements of such test waveforms, after passing through selected portions of the equipment or the complete system under evaluation, will give the required differential gain, phase and group delay information.

Stair-Step Generator

A modulated stair-step generator waveform is shown in Figure 1.9-64. It consists of five 20-IRE-unit risers with subcarrier modulation on each transition. The amplitude-linearity or differential gain response of an amplifier can be determined directly from oscilloscope measurements of the output wave display. By the use of a high pass filter the differential gain characteristic can be displayed more graphically (Figure 1.9-65, input), (Figure 1.9-66, output) showing appre-

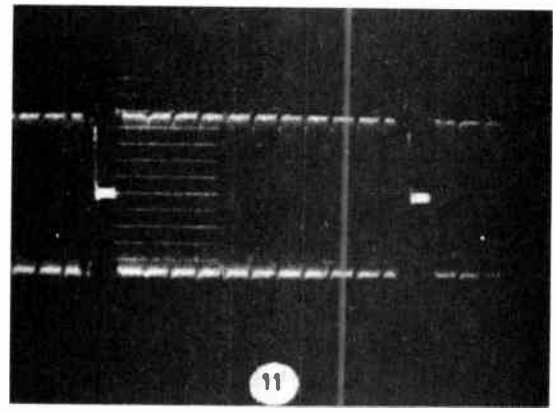


Figure 1.9-65. High pass filter output with modulated stair-step waveform input. (Picture courtesy of Marconi Instruments.)

ciable distortion. Differential phase measurements can be obtained by comparison of the subcarrier phase at each discrete level with phase of the color burst. Various oscillographic display techniques for precision phase measurements are available.

Sine-Squared Pulse and Bar

A second specialized waveform is the sine-squared pulse and bar with chrominance subcarrier modulation as shown in Figure 1.9-67. It evolved from the monochrome sine-squared pulse and bar shown in Figure 1.9-68. Use of this color test signal shows the presence of differential gain distortions as in Figure 1.9-69 and delay distortions as shown in Figure 1.9-70. Operationally the elegance of the method is in the direct display presentation where distortion limits may be checked by reticle overlay techniques.

Another frequently used waveform is the multiburst signal, Figure 1.9-71 and 72, which provides a series of selected frequency, constant-amplitude sine wave electrical bursts of 0.5 MHz, 1, 2, 3 and 4 sweep signals which sequentially sample all frequencies in

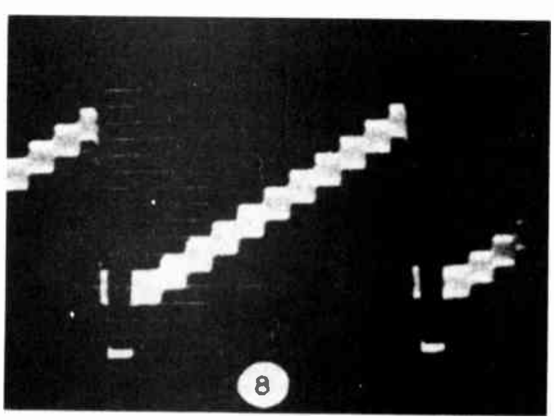


Figure 1.9-64. Modulated stair-step generator waveform. (Picture courtesy of Marconi Instruments.)

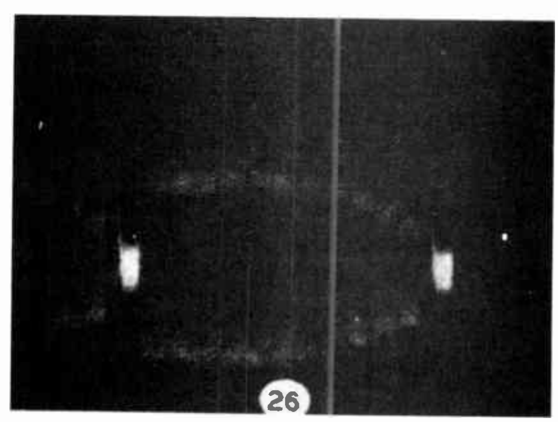


Figure 1.9-66. High pass filter output of modulated stair-step waveform showing large amount of differential gain error in amplifier under test. (Picture courtesy of Marconi Instruments.)

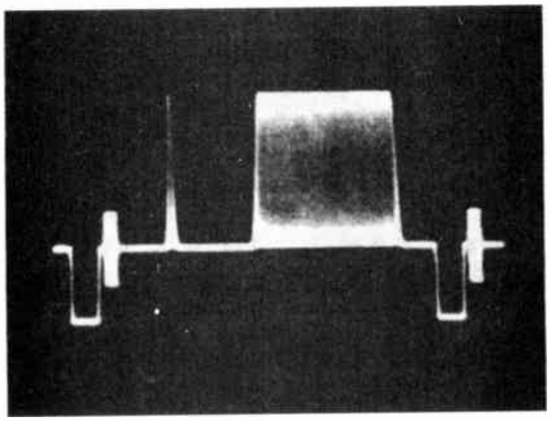


Figure 1.9-67. Combined luminance and chrominance sine-squared pulse and bar. (Picture courtesy of Marconi Instruments.)

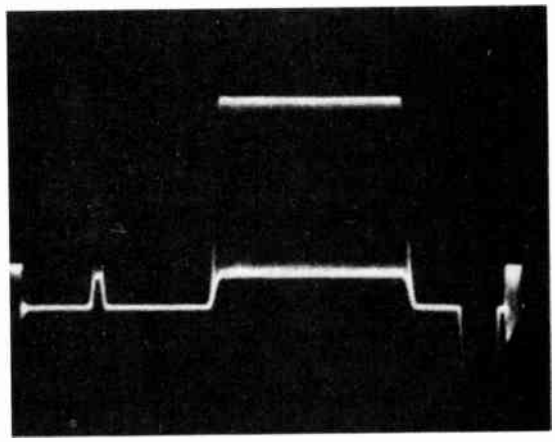


Figure 1.9-69. Gain inequality indicated by combined luminance and chrominance sine-squared pulse and bar. Compare with waveforms of Figure 1.9-67. (Picture courtesy of Marconi Instruments.)

the video pass band. However, it is more convenient to use and to interpret in routine frequency response tests of broadcast equipment.

Vectorscope

The *vectorscope* is a measurement instrument developed especially for color TV system tests and monitoring. Its essential feature is the polar or vectorial display of chrominance information in which the radial deflection is proportional to saturation of a color, and the angular position is equal to the phase angle of that color subcarrier with respect to the color burst. The 360° polar coordinate display corresponds to a complete cycle of color subcarrier or 280 nsec in a time display. By convention, the color burst is normalized at 180°. If the color bar signal described in Figures 1.9-61 and 63 is applied to the input to the vectorscope and the burst is normalized at 180°, the display shown in Figure 1.9-73 is obtained on the graticule.

It is noted that for standard signal levels each color vector in the color bar sequence falls within its approximately marked box on the graticule. The outer boxes define the FCC maximum permissible errors of ±10° in phase and ±20% in amplitude. The inner boxes correspond to ±2.5 phase error and 2.5% amplitude error.

A feature of the vectorscope color bar technique is that it gives immediate reassurance on system performance with a color bar test signal display.

By alternating two signal sources at the input, one can obtain direct readings on differential phase and amplitude behavior of any selected picture sources.

Vertical Interval Reference and Test Signals

The *vertical interval reference* (VIR) signal, consists of a chrominance bar having the same phase as color burst, together with an appropriate luminance pulse and a black level interval. The VIR signal is added to the main video signal and is in fact a certification that at the time it is added all conditions are normal. If various distortions occur to this VIR, it can

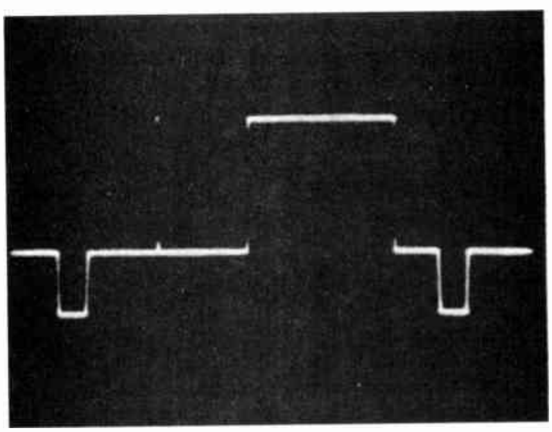


Figure 1.9-68. Monochrome sine-squared pulse and bar. (Picture courtesy of Marconi Instruments.)

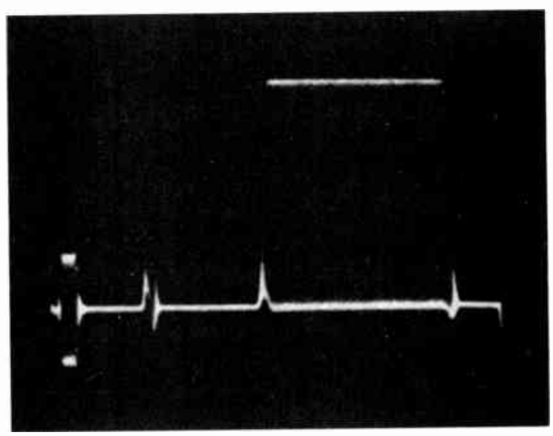


Figure 1.9-70. Delay inequality indicated by the combined luminance and chrominance sine-squared pulse and bar. (Picture courtesy of Marconi Instruments.)

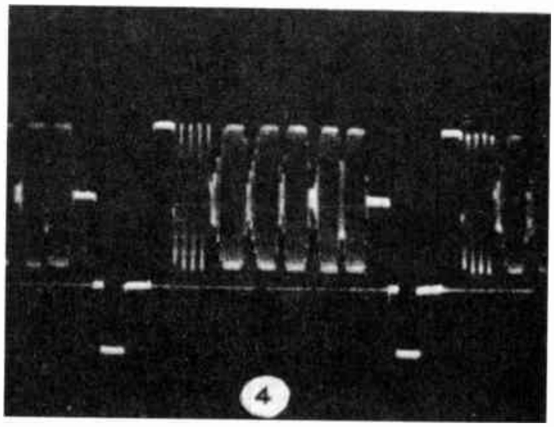


Figure 1.9-71. Multiburst test signal with burst at 0.5 MHz, 1, 2, 3 and 4 MHz. (Picture courtesy of Marconi Instruments.)

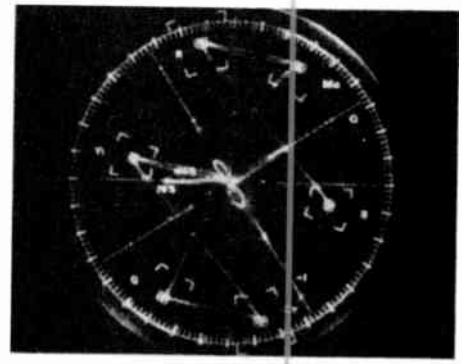


Figure 1.9-73. Vector display. Split field color bars 75% amplitude 100% white reference, 10% set up. Conforms to EIA specification RS 189. (Picture courtesy of Tektronix, Inc.)

be corrected, with the expectation that the main signal will also be corrected. Thus, more rigorous control and compensation of system errors is possible. A *vertical interval test* (VIT) signal is used to verify transmission conditions using multiburst, sine-squared or stair-step

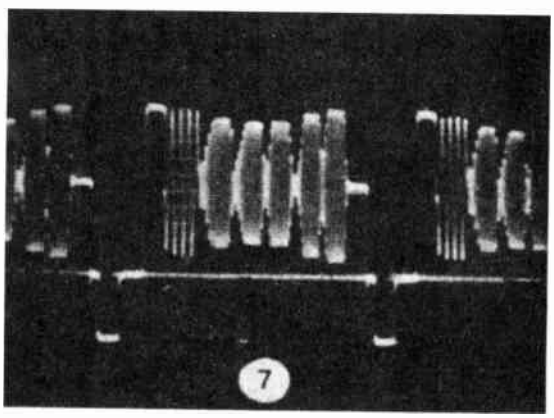


Figure 1.9-72. Multiburst output signal from amplifier having distortion. Compare with Figure 1.9-71. (Picture courtesy of Marconi Instruments.)

test signals. Such signals can be used for continuous monitoring of TV system performance.

A detailed discussion of video signal analysis and measurement can be found in Chapter 7.2, *Video Signal Analysis.*

REFERENCES

1. A detailed discussion of colorimetry and perception, and how these factors affect the viewer, is presented in "Color Television Engineering" by John W. Wentworth, McGraw-Hill Book Company, Inc., New York, 1955.
2. If the filters in the system are of minimum-phase type, only one of the plots is needed, for either plot can be derived from the other for this type of filter. Almost all common interstage coupling networks are of the minimum-phase type.
3. FM systems can have nonlinearity as a result of passive networks, but this is not considered here.
4. From the definition of differential gain by IRE Subcommittee 23.4.
5. From the definition of differential phase by IRE Subcommittee 23.4.
6. Tektronix Model 520 vectorscope is widely used for these measurements.

1.10
WORLDWIDE STANDARDS FOR CONVENTIONAL TELEVISION

D. H. PRITCHARD AND J. J. GIBSON

INTRODUCTION

The performance of a motion picture system in one location in the world is generally the same as in any other location. Thus, international exchange of film programming is relatively straightforward. This is not the case, however, with the conventional broadcast color television systems. The lack of compatibility has its origins in many factors, such as constraints in communications channel allocations and techniques, differences in local power source characteristics, network requirements, pickup and display technology, and political considerations relating to international telecommunications agreements.

Background

The most outstanding effort—as well as the most controversial effort—of the Eleventh Plenary Assembly of the International Radio Consultative Committee (CCIR), held in Oslo in 1966, was an attempt at standardization of color television systems by the participating countries of the world. The discussions pertaining to the possibility of a universal system proved inconclusive. Therefore, the International radio Consultative Committee, instead of issuing a unanimous recommendation for a single system CCIR, was forced to issue only a report describing the characteristics and recommendations for a variety of proposed systems. It was left to the controlling organizations of the individual countries to make their own choice as to which standard to adopt.

This outcome was not totally surprising because one of the primary requirements for any color television system is compatibility with a coexisting monochrome system. In many cases, the monochrome standards already existed and were dictated by such factors as local power line frequencies (relevant to field and frame rates) as well as radio frequency channel allocations and pertinent telecommunications agreements.

Thus, such technical factors as line number, field rate, video bandwidth, modulation technique, and sound carrier frequencies were predetermined and varied in many regions of the world. The ease with which international exchange of program material may be accomplished is thereby hampered, and has been accommodated over the years by means of standards conversion techniques, or *transcoders*, with varying degrees of loss in quality. While invariably introducing some compromises, standards conversion techniques have provided surprisingly good service in recent years with the growing use of satellite relays coupled with the advances in digital signal processing in both the video and audio domains.

MONOCHROME COMPATIBLE COLOR TV SYSTEMS

In order to achieve success in the introduction of a color television system, it is essential that the color system be fully compatible with the existing black-and-white system. That is, monochrome receivers must be able to produce high-quality black-and-white images from a color broadcast, and color receivers must produce high-quality black-and-white images from monochrome broadcasts. The first such color television system to be placed into commercial broadcast service was developed in the United States. On 17, December 1953, the Federal Communications Commission (FCC) approved transmission standards and authorized broadcasters, as of January 23, 1954, to provide regular service to the public under these standards. This decision was the culmination of the work of the National Television System Committee (NTSC) upon whose recommendation the FCC action was based[1]. Subsequently, this system, commonly referred to as the NTSC system, was adopted by Canada, Japan, Mexico, and others.

That more than 40 years ago, these standards are still providing color television service of good quality testifies to the validity and applicability of the fundamental principles underlying the choice of specific techniques and numerical standards.

The previous existence of a monochrome television standard was two-edged in that it provided a foundation upon which to build the necessary innovative techniques while simultaneously imposing the requirement of compatibility. Within this framework, an underlying theme—that which the eye does not see does not need to be transmitted nor reproduced—set the stage for a variety of fascinating developments in what has been characterized as an "economy of representation"[1].

The countries of Europe delayed the adoption of a color television system, and in the years between 1953 and 1967, a number of alternative systems that were compatible with the 625-line, 50-field existing monochrome systems were devised. The development of these systems was to some extent influenced by the fact that the technology necessary to implement some of the NTSC requirements was still in its infancy. Thus, many of the differences between NTSC and the other systems are the result of technological rather than fundamental theoretical considerations.

Most of the basic techniques of NTSC are incorporated into the other system approaches. For example, the use of wideband luminance and relatively narrowband chrominance, following the teachings of the principle of *mixed highs*, is involved in all systems. Similarly, the concept of providing horizontal interlace for reducing the visibility of the color subcarrier(s) is followed in all approaches. This feature is required to reduce the visibility of signals carrying color information that are contained within the same frequency range as the coexisting monochrome signal, thus maintaining a high order of compatibility.

An early system that received approval was one proposed by Henri de France of the Compagnie de Television of Paris. It was argued that if color could be relatively band-limited in the horizontal direction, it could also be band-limited in the vertical direction. Thus, the two pieces of coloring information (hue and saturation) that need to be added to the one piece of monochrome information (brightness) could be transmitted as subcarrier modulation that is sequentially transmitted on alternate lines—thereby avoiding the possibility of unwanted crosstalk between color signal components. Thus, at the receiver, a one-line memory, commonly referred to as a 1-*H* delay element, must be employed to store one line to then be concurrent with the following line. Then a linear matrix of the red and blue signal components (R and B) is used to produce the third green component (G). Of course, this necessitates the addition of a line-switching identification technique. Such an approach, designated as *S*equential *C*ouleur *A*vec *M*emoire, for sequential color with memory (SECAM) was developed and officially adopted by France and the USSR, and broadcast service began in France in 1967.

The implementation technique of a 1-*H* delay element led to the development, largely through the efforts of Walter Bruch of Telefunken Company, of the *Phase Alternation Line* (PAL) system. This approach was aimed at overcoming an implementation problem of NTSC that requires a high order of phase and amplitude integrity (skew-symmetry) of the transmission path characteristics about the color subcarrier to prevent color quadrature distortion. The line-by-line alternation of the phase of one of the color signal components averages any colorimetric distortions to the observer's eye to that of the correct value. The system in its simplest form (simple PAL), however, results in line flicker (Hanover bars). The use of a 1-*H* delay device in the receiver greatly alleviates this problem (standard PAL). PAL systems also require a line identification technique.

The standard PAL system was adopted by numerous countries in continental Europe, as well as in the United Kingdom. Public broadcasting began in 1967 in Germany and the United Kingdom using two slightly different variants of the PAL system.

NTSC, PAL, AND SECAM SYSTEMS OVERVIEW

In order to properly understand the similarities and differences of the conventional television systems, a familiarization with the basic principles of NTSC, PAL, and SECAM is required. As previously stated, because many basic techniques of NTSC are involved in PAL and SECAM, a thorough knowledge of NTSC is necessary in order to understand PAL and SECAM.

The same R, G, and B pickup devices and three primary color display devices are used in all systems. The basic camera function is to analyze the spectral distribution of the light from the scene in terms of its red, green, and blue components on a point-by-point basis as determined by the scanning rates. The three resulting electrical signals must then be transmitted over a band-limited communications channel to control the three-color display device to make the *perceived* color at the receiver appear essentially the same as the *perceived* color at the scene.

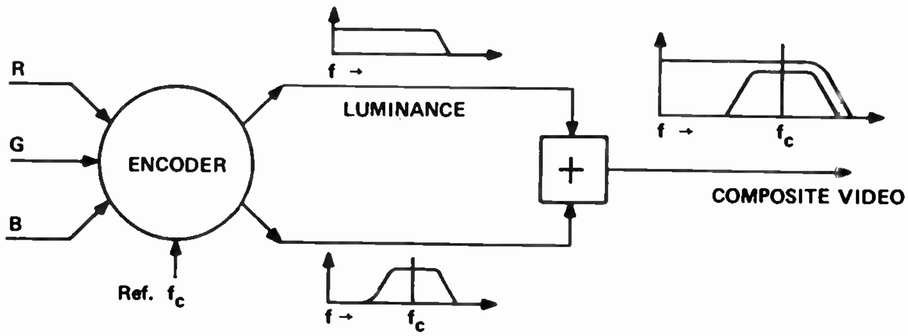
It is useful to define color as a psycho-physical property of light—specifically, as the combination of those characteristics of light that produce the sensations of brightness, hue, and saturation. Brightness refers to the relative intensity; hue refers to that attribute of color that allows separation into spectral groups perceived as red, green, yellow, and so on (in scientific terms, the *dominant wavelength*); and saturation is the degree to which a color deviates from a neutral gray of the same brightness—the degree to which it is "pure," or "pastel," or "vivid." These three characteristics represent the total information necessary to define and/or recreate a specific color stimulus.

This concept is useful to communication engineers in developing encoding and decoding techniques to efficiently compress the required information within a given channel bandwidth and to subsequently recombine the specific color signal values in the proper proportions at the reproducer. The NTSC color standards define the first commercially broadcast process for achieving this result.

A preferred signal arrangement was developed that resulted in reciprocal compatibility with monochrome pictures and was transmitted within the existing monochrome channel, as shown in Figure 1.10-1. Thus, one signal (luminance) is chosen in all approaches to occupy the wide-band portion of the channel and to convey the *brightness* as well as the detail information content. A second signal (chrominance), representative of the chromatic attributes of *hue* and *saturation*, is assigned less channel width in accordance with the principle that, in human vision, full three-color reproduction is not required over the entire range of resolution—commonly referred to as the *mixed-highs principle*.

Another fundamental principle employed in all systems involves arranging the chrominance and luminance signals within the same frequency band without excessive mutual interference. Recognition that the

- COMPATIBILITY WITH CO-EXISTING MONOCHROME SYSTEM.
- ENCODE WIDEBAND R, G, B COLOR PRIMARY SIGNALS.
  - WIDEBAND LUMINANCE (BRIGHTNESS)
  - NARROW-BAND MODULATION OF A COLOR SUBCARRIER (Hue and Saturation)



- SUBCARRIER FREQUENCY INTERLACE
  - ODD MULTIPLE OF ½H TO REDUCE VISIBILITY OF CHROMINANCE INFORMATION SUBCARRIER.

Figure 1.10-1. Basic communications channel principles applied to color television.

scanning process, being equivalent to sampled-data techniques, produces signal components largely concentrated in uniformly spaced groups across the channel width, led to introduction of the concept of horizontal frequency interlace (dot interlace). The color subcarrier frequency is so chosen as to be an odd multiple of one-half the line rate (in the case of NTSC) such that the phase of the subcarrier is exactly opposite on successive scanning lines. This substantially reduces the subjective visibility of the color signal "dot" pattern components.

Thus, the major differences among the three main systems of NTSC, PAL, and SECAM are in the specific modulating processes used for encoding and transmitting the chrominance information. The similarities and differences are briefly summarized in Table 1.10-1.

- ALL SYSTEMS:
  - THREE-PRIMARY ADDITIVE COLORIMETRIC PRINCIPLES
  - SIMILAR CAMERA PICK-UP AND RECEIVER DISPLAY TECHNOLOGY
  - WIDEBAND LUMINANCE AND NARROW-BAND COLOR
- COMPATIBILITY WITH CO-EXISTING MONOCHROME SYSTEM:

  INTRODUCES FIRST ORDER DIFFERENCES
  - LINE NUMBER
  - FIELD/FRAME RATES
  - BANDWIDTH
  - FREQUENCY ALLOCATION
- MAJOR DIFFERENCES IN COLOR TECHNIQUES
  - NTSC – PHASE AND AMPLITUDE QUADRATURE MODULATION OF INTERLACED SUBCARRIER
  - PAL – SIMILAR TO NTSC BUT WITH LINE ALTERNATION OF "V" COMPONENT
  - SECAM – FREQUENCY MODULATION OF LINE SEQUENTIAL COLOR SUBCARRIERS

Table 1.10-1. General comparison of worldwide television systems.

The NTSC Color System

The importance of the colorimetric concepts of brightness, hue, and saturation comprising the three pieces of information necessary to analyze or recreate a specific color value becomes evident in the formation of the composite color television NTSC format.

The luminance, or monochrome, signal is formed by addition of specific proportions of the red, green, and blue signals and occupies the total available video bandwidth of 0–4.2 MHz. The NTSC, PAL, and SECAM systems all use the same luminance (Y) signal formation, differing only in the available bandwidths.

The Y signal components have relative voltage values representative of the brightness sensation in the human eye. Therefore, the red, green, and blue voltage components are tailored in proportion to the standard luminosity curve at the particular values of the dominant wavelengths of the three color primaries chosen for color television. Thus, the luminance signal makeup for all systems, as normalized to white, is described by:

$$E'_Y = 0.299E'_R + 0.587E'_G + 0.114E'_B \qquad (1)$$

The signal of Equation (1) would be exactly equal to the output of a linear monochrome sensor with ideal spectral sensitivity if the red, green, and blue elements were also linear devices with theoretically correct spectral-sensitivity curves. In actual practice, the red, green, and primary signals are deliberately made nonlinear to accomplish *gamma correction* (adjustment of the slope of the input/output transfer characteristic). The prime mark (′) is used to denote a gamma-corrected signal.

Table 1.10-2 gives the equations for the chrominance signal components. Signals representative of the chromaticity information (hue and saturation) that relate to the differences between the luminance signal and the basic red, green, and blue signals are generated in a linear matrix. This new set of signals is termed *color-difference* signals and is designated as R − Y, G − Y, and B − Y. These signals modulate a subcarrier that is combined with the luminance component and passed through a common communications channel. At the receiver, the color difference signals are detected, separated and individually added to the luminance signal in three separate paths to recreate the original R, G and B signals according to the equations:

$$E'_Y + E'_{(R-Y)} = E'_Y + E'_R - E'_Y = E'_R \qquad (2a)$$

$$E'_Y + E'_{(G-Y)} = E'_Y + E'_G - E'_Y = E'_G \qquad (2b)$$

$$E'_Y + E'_{(B-Y)} = E'_Y + E'_B - E'_Y = E'_B \qquad (3c)$$

In the specific case of NTSC two other color-difference signals, designated as *I* and *Q*, are formed at the encoder and are used to modulate the color subcarrier.

Another reason for the choice of signal values in the NTSC system is that the eye is more responsive to spatial and temporal variations in luminance than it is to variations in chrominance. Therefore, the visibility of luminosity changes resulting from random noise and interference effects may be reduced by properly proportioning the relative chrominance gain and encoding angle values with respect to the luminance values. Thus, the *principle of constant luminance* is incorporated into the system standard.[1,2]

LUMINANCE:

$$E'_Y = 0.299\ E'_R + 0.587\ E'_G + 0.114\ E'_B$$

(Common for all systems)

CHROMINANCE:

NTSC

$$E'_I = -0.274\ E'_G + 0.596\ E'_R - 0.322\ E'_B$$

$$E'_Q = -0.522\ E'_G + 0.211\ E'_R + 0.311\ E'_B$$

$$B\text{-}Y = 0.493\ (E'_B - E'_Y)$$

$$R\text{-}Y = 0.877\ (E'_R - E'_Y)$$

$$G\text{-}Y = 1.413\ (E'_G - E'_Y)$$

PAL

$$E'_U = 0.493\ (E'_B - E'_Y)$$

$$\pm E'_V = \pm 0.877\ (E'_R - E'_Y)$$

SECAM

$$D'_R = -1.9\ (E'_R - E'_Y)$$

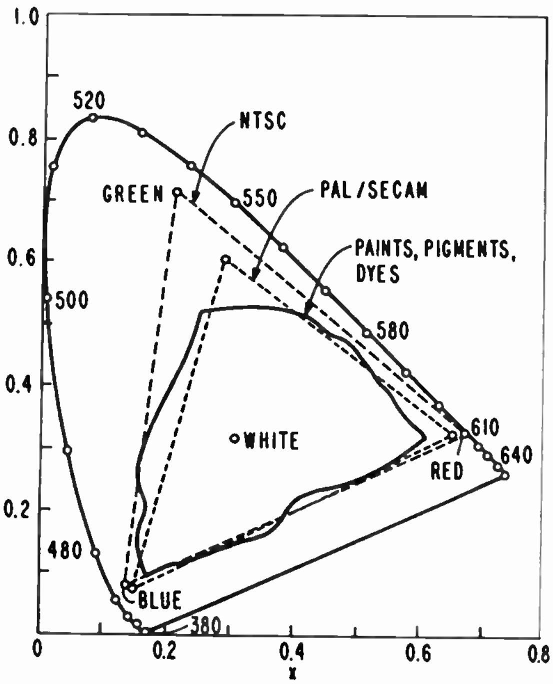
$$D'_B = 1.5\ (E'_B - E'_Y)$$

Table 1.10-2. Electronic color signal values for NTSC, PAL, and SECAM.

The voltage outputs from the three camera sensors are adjusted to be equal when a scene reference white or neutral gray object is being scanned for the color temperature of the scene ambient. Under this condition, the color subcarrier also automatically becomes zero. The colormetric values have been formulated by assuming that the reproducer will be adjusted for *illuminant C*, representing the color of average daylight.

Figure 1.10-2 is a CIE chromaticity diagram indicating the primary color coordinates for NTSC, PAL, and SECAM. It is interesting to compare the available color gamut relative to that of all color paint, pigment, film, and dye processes.

In the NTSC color standard, the chrominance information is carried as simultaneous amplitude and phase modulation of a subcarrier chosen to be in the high frequency portion of the 0–4.2 MHz video band and specifically related to the scanning rates as an odd multiple of one-half the horizontal line rate, as shown by the vector diagram in Figure 1.10-3. The hue information is assigned to the instantaneous phase of the subcarrier. Saturation is determined by the *ratio* of the



		x	y
NTSC	R =	0.67	0.33
	G =	0.21	0.71
	B =	0.14	0.08
PAL/SECAM	R =	0.64	0.33
	G =	0.29	0.60
	B =	0.15	0.06
WHITE:	NTSC(ILL. C) =	0.310	0.316
	PAL/SECAM(D6500) =	0.313	0.329

Figure 1.10-2. CIE chromacity diagram-system comparison.

instantaneous amplitude of the subcarrier to that of the corresponding luminance signal amplitude value.

The choice of the *I* and *Q* color modulation components relates to the variation of color acuity characteristics of human color vision as a function of the field of view and spatial dimensions of objects in the scene. The color acuity of the eye decreases as the size of the viewed object is decreased and thereby occupies a small part of the field of view. Small objects, represented by frequencies above about 1.5–2.0 MHz, produce no color sensation (*mixed-highs*). Intermediate spatial dimensions (approximately in the 0.5–1.5 MHz range) are viewed satisfactorily if reproduced along a preferred orange-cyan axis. Large objects (0–0.5 MHz) require full three-color reproduction for subjectively-pleasing results. Thus, the *I* and *Q* bandwidths are chosen accordingly, and the preferred colorimetric reproduction axis is obtained when only the *I* signal exists by rotating the subcarrier modulation vectors by 33. In this way, the principles of mixed-highs and *I*, *Q color-acuity axis* operation are exploited.

At the encoder, the *Q* signal component is band-limited to about 0.6 MHz and is representative of the green-purple color-axis information. The *I* signal component has a bandwidth of about 1.5 MHz and contains the orange-cyan color axis information. These two signals are then used to individually modulate the color subcarrier in two balanced modulators operated in phase quadrature. The *sum products* are selected and added to form the composite chromaticity subcarrier. This signal—in turn—is added to the luminance signal along with the appropriate horizontal and vertical synchronizing and blanking waveforms to include the color-synchronization burst. The result is the total composite color video signal.

Quadrature synchronous detection is used at the receiver to identify the individual color signal components. When individually recombined with the luminance signal, the desired R, G and B signals are recreated. The receiver designer is free to demodulate either at *I* or *Q* and matrix to form B − Y, R − Y and G − Y, or as in nearly all modern receivers, at B − Y and R − Y and maintain 500 kHz equiband color signals.

The chrominance information can be carried without loss of identity provided that the proper phase relationship is maintained between the encoding and decoding processes. This is accomplished by transmitting a reference burst signal consisting of eight or nine cycles of the subcarrier frequency at a specific phase [−(B − Y)] following each horizontal synchronizing pulse as shown in Figure 1.10-4.

The specific choice of color subcarrier frequency in NTSC was dictated by at least two major factors. First, the necessity for providing horizontal interlace in order to reduce the visibility of the subcarrier requires that the frequency of the subcarrier be precisely an odd multiple of one-half the horizontal line rate. Figure 1.10-5 shows the energy spectrum of the composite NTSC signal for a typical stationary scene. This interlace provides line-to-line phase reversal of the color subcarrier, thereby reducing its visibility (and thus improving compatibility with monochrome reception). Second, it is advantageous to also provide interlace of the beat-frequency (about 920 kHz) occurring between the color subcarrier and the average value of the sound carrier. For total compatibility reasons, the sound carrier was left unchanged at 4.5 MHz and the line number remained at 525. Thus, the resulting line scanning rate and field rate varied slightly from that of the monochrome values, but stayed within the previously existing tolerances. The difference is exactly one part in one thousand. Specifically, the line rate is 15.734 kHz, the field rate is 59.94 Hz, and the color subcarrier is 3.578545 MHz.

PAL Color System

Except for some minor details, the color encoding principles for PAL are the same as those for NTSC. However, the phase of the color signal, $E_V = R - Y$, is reversed by 180° from line-to-line. This is done for the purpose of averaging, or canceling, certain color errors resulting from amplitude and phase distortion of the color modulation sidebands. Such distortions

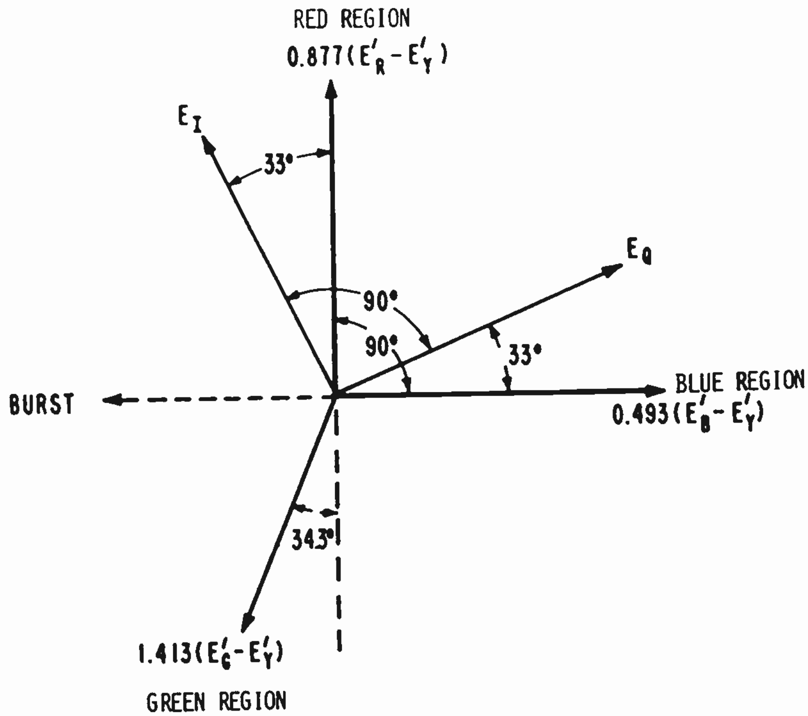


Figure 1.10-3. NTSC color modulation phase diagram.

might occur as a result of equipment or transmission path problems.

The NTSC chroma signal expression within the frequency band common to both I and Q is given by:

$$C_{NTSC} = \frac{B - Y}{2.03} \sin \omega_{SC}t + \frac{R - Y}{1.14} \cos \omega_{SC}t \quad (3)$$

The PAL chroma signal expression is given by:

$$C_{PAL} = \frac{U}{2.03} \sin \omega_{SC}t \pm \frac{C}{1.14} \cos \omega_{SC}t \quad (4)$$

where U and $\pm V$ have been substituted for $B - Y$ and $R - Y$ signal values, respectively.

The PAL format employs equal bandwidths for the U and V color-difference signal components that are about the same as the NTSC I signal bandwidth (1.3 MHz at 3 dB). There are slight differences in the U and V bandwidth in different PAL systems because of the differences in luminance bandwidth and sound carrier frequencies. (See the applicable ITU-R documents for specific details.)

The V component was chosen for the line-by-line reversal process because it has a lower gain factor than U and, therefore, is less susceptible to switching rate ($\frac{1}{2} f_H$) imbalance. Figure 1.10-6 indicates the vector diagram for the PAL quadrature modulated and line-alternating color modulation approach.

The result of the switching of the V signal phase at the line rate is that any phase errors produce complementary errors from V into the U channel. In addition, a corresponding switch of the decoder V channel results in a constant V component with complementary errors from the U channel. Thus, any line-to-line averaging process at the decoder, such as the retentivity of the eye (simple PAL) or an electronic averaging technique such as the use of a 1-H delay element (standard PAL), produces cancellation of the phase (hue) error and provides the correct hue but with somewhat reduced saturation—this error being subjectively much less visible.

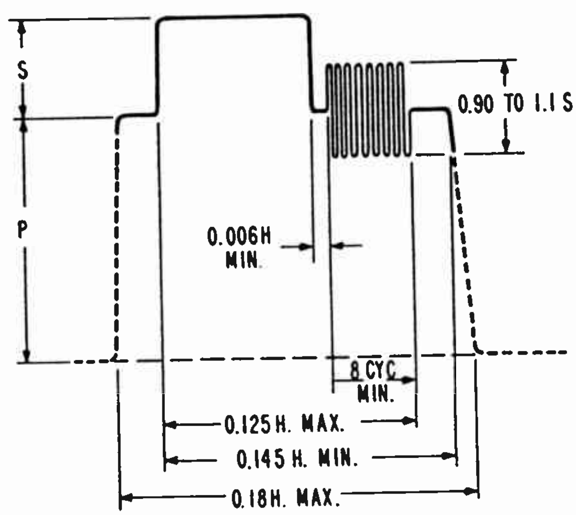


Figure 1.10-4. NTSC color burst synchronizing signal.

Obviously, the PAL receiver must be provided with some means by which the V signal switching sequence

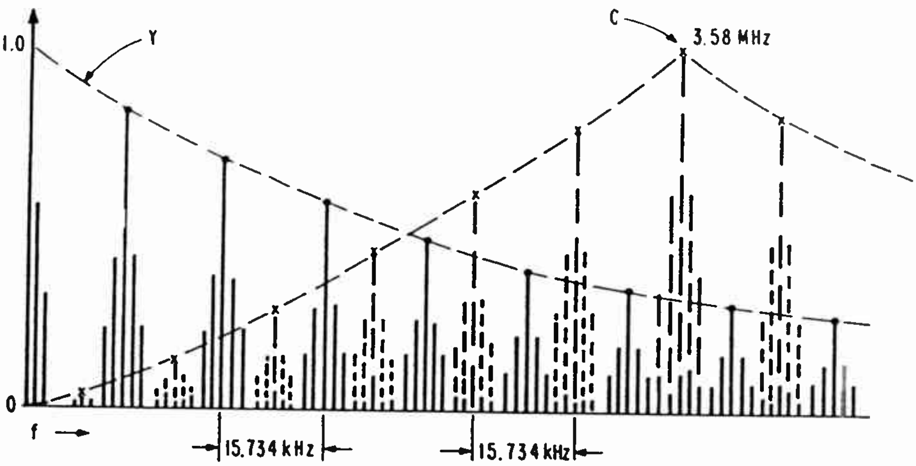


Figure 1.10-5. Luminance/chrominance horizontal frequency interface principle.

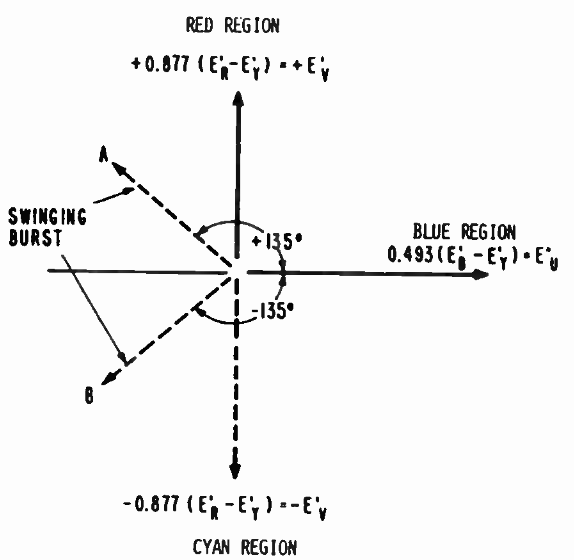


Figure 1.10-6. PAL color modulation phase diagram.

may be identified. The technique employed is known as *A B* sync, PAL sync, or *swinging burst* and consists of alternating the phase of the reference burst by ±45° at a line rate as shown in Figure 1.10-6. The burst is constituted from a fixed value of *U* phase and a switched value of *V* phase. Because the sign of the *V* burst component is the same sign as the *V* picture content, the necessary switching "sense" or identification information is available. At the same time, the fixed-*U* component is used for reference carrier synchronization.

Figure 1.10-7 explains the degree to which horizontal frequency (dot) interlace of the color subcarrier components with the luminance components is achieved in PAL and may be summarized as follows: in NTSC, the *Y* components are spaced at f_H intervals as a result of the horizontal sampling (blanking) process. Thus, the choice of a color subcarrier whose harmonics are also separated from each other by f_H (as they are odd multiples of $\frac{1}{2} f_H$) provides a half-line offset and results in a perfect dot interlace pattern that moves upward. Four complete field scans are required to repeat a specific picture element dot position.

In PAL, the luminance components are also spaced at f_H intervals. Because the *V* components are switched symmetrically at half the line rate, only odd harmonics exist, with the result that the *V* components are spaced at intervals of f_H. They are spaced at half-line intervals from the *U* components which, in turn have f_H spacing intervals due to blanking. If half-line offset were used,

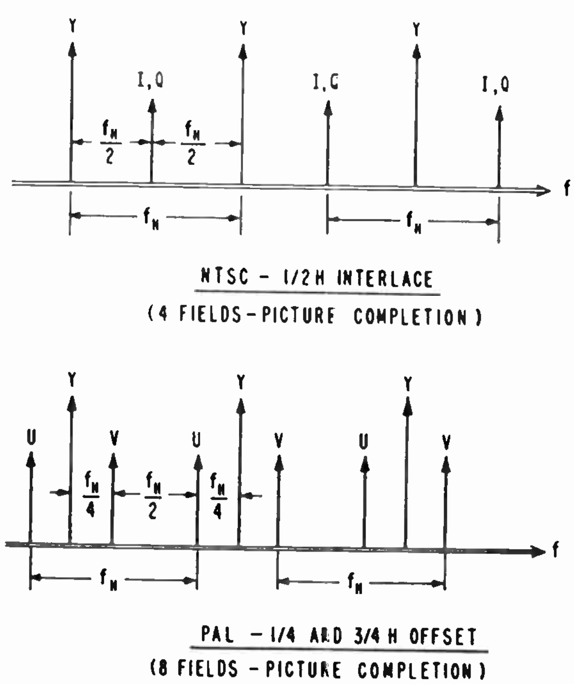


Figure 1.10-7. NTSC and PAL frequency interface relationship.

the U components would be perfectly interlaced but the V components would coincide with Y and, thus, not be interlaced, creating vertical, stationary dot patterns.

For this reason, in PAL, a $\frac{1}{4}$-line offset for the subcarrier frequency is used as shown in Figure 1.10-7. The expression for determining the PAL subcarrier specific frequency for 625-line/50-field systems is given by

$$F_{SC} = \frac{1135}{4} f_H + \frac{1}{2} f_V \qquad (5)$$

The additional factor $\frac{1}{2} f_V = 25$ Hz is introduced to provide motion to the color dot pattern, thereby reducing its visibility. The degree to which interlace is achieved is, therefore, not perfect, but is acceptable, and eight complete field scans must occur before a specific picture element dot position is repeated.

One additional function must be accomplished in relation to PAL color synchronization. In all systems, the burst signal is eliminated during the vertical synchronization pulse period. Because, in the case of PAL, the swinging burst phase is alternating line-by-line, some means must be provided for ensuring that the phase is the same for the first burst following vertical sync on a field-by-field basis. Therefore, the burst reinsertion time is shifted by one line at the vertical field rate by a pulse referred to as the *meander* gate. The timing of this pulse relative to the A versus B burst phase is shown in Figure 1.10-8.

The transmitted signal specifications for PAL systems include the basic features discussed previously. Although a description of a great variety of receiver decoding techniques is outside the scope and intent of this chapter, it is appropriate to review—at least briefly—the following major features: "Simple" PAL relies upon the eye to average the line-by-line color switching process and can be plagued with line beats known as *Hanover bars* caused by the system nonlinearities introducing visible luminance changes at the line rate. "Standard" PAL employs a 1-H delay line element to separate U color signal components from V color signal components in an averaging technique, coupled with summation and subtraction functions. Hanover bars can also occur in this approach if an imbalance of amplitude or phase occurs between the delayed and direct paths.

In a PAL system, vertical resolution in chrominance is reduced as a result of the line averaging processes. The visibility of the reduced vertical color resolution as well as the vertical time coincidence of luminance and chrominance transitions differs depending upon whether the total system, transmitter through receiver, includes one or more averaging (comb filter) processes.

Thus, PAL provides a similar system to NTSC and has gained favor in many areas of the world, particularly for 625-line/50-field systems.

SECAM Color System

The optimized SECAM system, known as SECAM III, is the system adopted by France and the USSR in 1967. The SECAM method has several features in common with NTSC such as the same E'_Y signal and the same $E'_B - E'_Y$ and $E'_R - E'_Y$ color-difference signals. However, this approach differs considerably from NTSC and PAL in the manner in which the color information is modulated onto the subcarrier(s).

First, the $R - Y$ and $B - Y$ color difference signals are transmitted alternately in time sequence from one successive line to the next, the luminance signal being

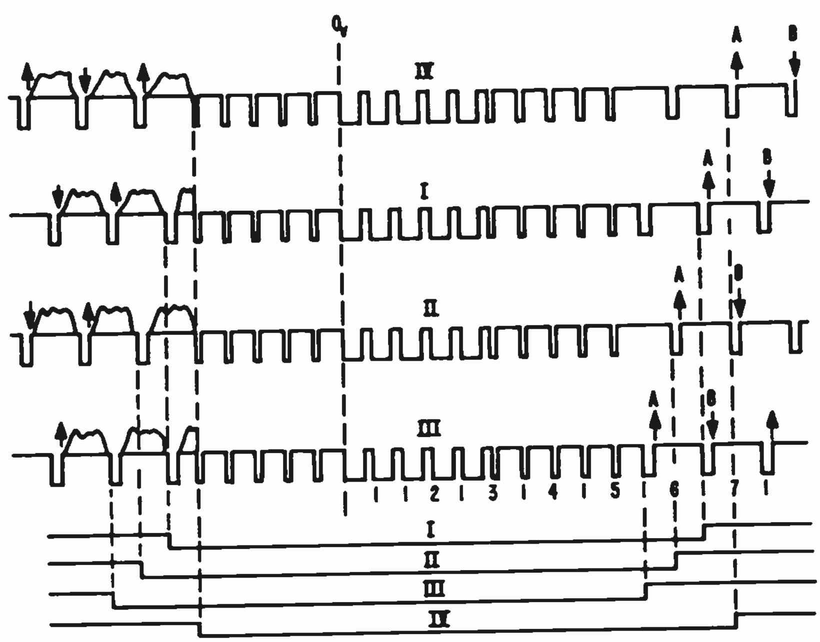


Figure 1.10-8. PAL "meander" burst blanking gate timing diagram for B, G, H, and I Pal.

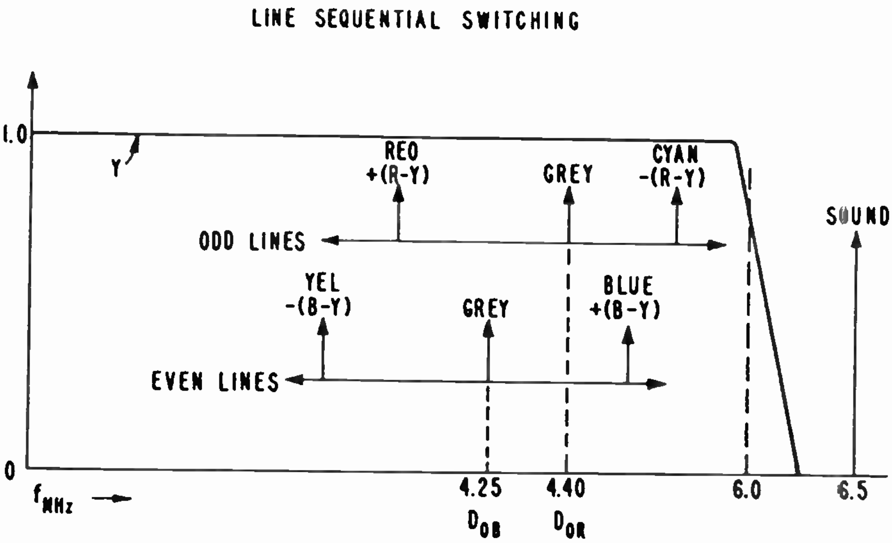


Figure 1.10-9. SECAM FM color modulation system.

common to every line. Because there is an odd number of lines, any given line carries $R - Y$ information on one field and $B - Y$ information on the next field. Second, the $R - Y$ and $B - Y$ color information is conveyed by frequency modulation of different subcarriers. Thus, at the decoder, a 1-H delay element, switched in time synchronization with the line switching process at the encoder, is required in order to have simultaneous existence of the $B - Y$ and $R - Y$ signals in a linear matrix to form the $G - Y$ component.

The $R - Y$ signal is designated as D'_R and the $B - Y$ signal as D'_B. The undeviated frequency for the two subcarriers, respectively, is determined by

$$\begin{aligned} F_{OB} &= 272\, f_H = 4.250000 \text{ MHz} \\ F_{OR} &= 282\, f_H = 4.406250 \text{ MHz} \end{aligned} \qquad (6)$$

These frequencies represent zero color difference information (zero output from the FM discriminator), or a neutral gray object in the televised scene.

As shown in Figure 1.10-9, the accepted convention for direction of frequency change with respect to the polarity of the color difference signal is opposite for the D_{OB} and D_{OR} signals. A positive value of D_{OR} means a decrease in frequency, whereas a positive value of D_{OB} indicates an increase in frequency. This choice relates to the idea of keeping the frequencies representative of the most critical color away from the upper edge of the available bandwidth to minimize the instrumentation distortions.

The deviation for D'_R is 280 kHz and D'_B is 230 kHz. The maximum allowable deviation, including preemphasis, for $D'_R = -506$ kHz and $+350$ kHz, while the values for $D'_B = -350$ kHz and $+506$ kHz.

Two types of preemphasis are employed simultaneously in SECAM. First, as shown in Figure 1.10-10, a conventional type of preemphasis of the low-frequency color difference signals is introduced. The characteristic is specified to have a reference level break-point at 85 kHz (f_1) and a maximum emphasis of 2.56 dB. The expression for the characteristic is given as

$$A = \frac{1 + j(f/f_1)}{1 + J(f/3f_1)} \qquad (7)$$

A second form of preemphasis (Figure 1.10-10) is introduced at the subcarrier level where the amplitude

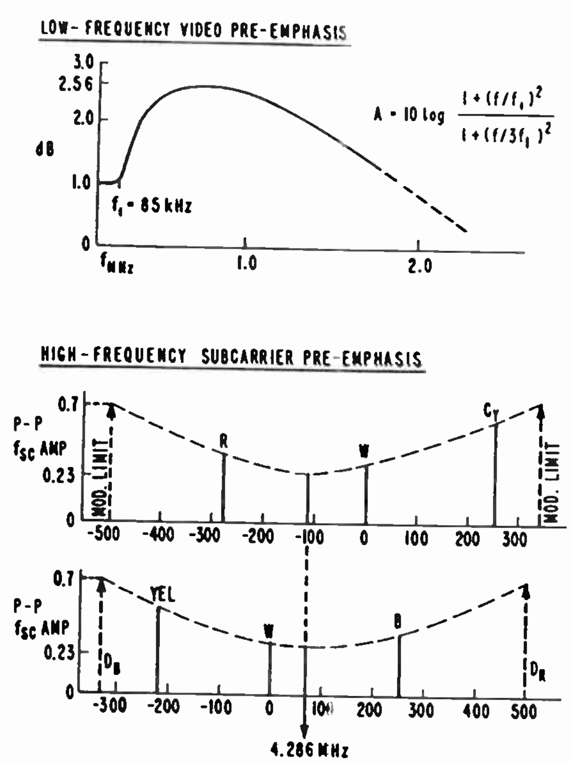


Figure 1.10-10. SECAM color signal pre-emphasis.

FIELD	LINE #	COLOR	SUBCARRIER Ø
ODD (1)	n	f_{OR}	0°
EVEN (2)	n + 313	f_{OB}	180°
ODD (3)	n + 1	f_{OB}	0°
EVEN (4)	n + 314	f_{OR}	0°
ODD (5)	n + 2	f_{OR}	180°
EVEN (6)	n + 315	f_{OB}	180°
ODD (7)	n + 3	f_{OB}	0°
EVEN (8)	n + 316	f_{OR}	180°
ODD (9)	n + 4	f_{OR}	0°
EVEN (10)	n + 317	f_{OB}	0°
ODD (11)	n + 5	f_{OB}	180°
EVEN (12)	n + 318	f_{OR}	180°

Note:
- 2 frames (4 fields) for picture completion.
- Subcarrier interlace is field-to-field and line-to-line of same color.

Figure 1.10-11. Color vs. line and field timing relationship for SECAM.

of the subcarrier is changed as a function of the frequency deviation. The expression for this inverted bell shaped characteristic is given as

$$G = M_O \frac{1 \times j16\left(\frac{f}{f_c} - \frac{f_c}{f}\right)}{1 + j1.26\left(\frac{f}{f_c} - \frac{f_c}{f}\right)} \qquad (8)$$

where: f = 4.286 MHz and 2M = 23 percent of the luminance amplitude (100 IRE).

This type of preemphasis is intended to further reduce the visibility of the frequency modulated subcarriers in low luminance level color values and to improve the signal-to-noise ratio (SNR) in high luminance and highly saturated colors. Thus, monochrome compatibility is better for pastel average picture level objects but sacrificed somewhat in favor of SNR in saturated color areas.

Of course, precise interlace of frequency modulated subcarriers for all values of color modulation cannot occur. Nevertheless, the visibility of the interference represented by the existence of the subcarriers may be reduced somewhat by the use of two separate carriers, as is done in SECAM. Figure 1.10-11 indicates the line-switching sequence in that at the undeviated "resting" frequency situation, the two-to-one vertical interlace in relation to the continuous color difference line-switching sequence produces adjacent line pairs of f_{OB} and f_{OR} signals. In order to further reduce the subcarrier dot visibility, the phase of the subcarriers (phase carries no picture information in this case) is reversed 180 degrees on every third line and between each field. This, coupled with the bell preemphasis, produces a degree of monochrome compatibility considered subjectively adequate.

As in PAL, the SECAM system must provide some means for identifying the line-switching sequence between the encoding and decoding processes. This is accomplished by introducing alternate D_R and D_B color identifying signals for nine lines during the vertical blanking interval following the equalizing pulses after vertical sync (see Figure 1.10-12). These bottle shaped signals occupy a full line each and represent the frequency deviation in each time sequence of D_B and D_R at zero luminance value. These signals can be thought of as fictitious green color that is used at the decoder to determine the line-switching sequence.

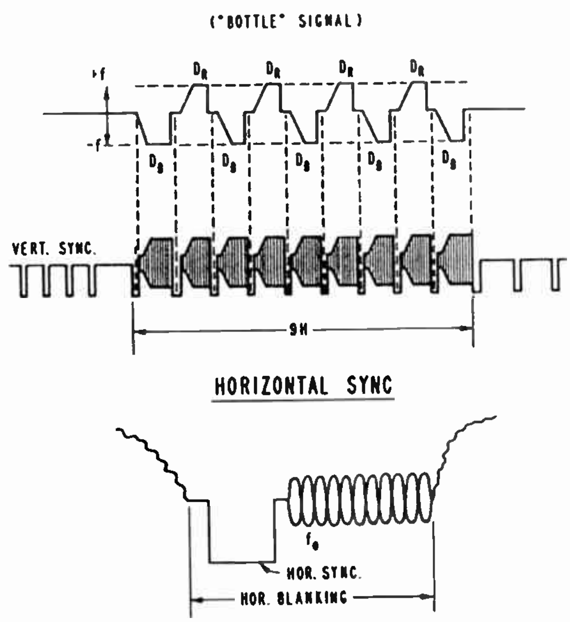


Figure 1.10-12. SECAM line identification signal.

During horizontal blanking, the subcarriers are blanked and a burst of f_{OB}/f_{OR} is inserted and used as a gray level reference for the FM discriminators, to establish their proper operation at the beginning of each line.

Thus, the SECAM system is a line sequential color approach using frequency modulated subcarriers. A special identification signal is provided to identify the line-switch sequence and is especially adapted to the 625-line/50-field wideband systems available in France and the USSR.

It should be noted that SECAM, as practiced, employs amplitude modulation of the sound carrier as opposed to the FM sound modulation in other systems.

Additional Systems of Historical Interest

Of the numerous system variations proposed over the intervening years since the development of the NTSC system, at least two others, in addition to PAL and SECAM, should be mentioned briefly. The first of these was additional reference transmission (ART), which involved the transmission of a continuous reference pilot carrier in conjunction with a conventional NTSC color subcarrier quadrature modulation signal. A modification of this scheme involved a *multiburst* approach that utilized three color bursts, one at black level, one at intermediate gray level, and one at white level, to be used for correcting differential phase distortion.

Another system, perhaps better known, was referred to as NIR or SECAM IV. Developed in the USSR Nauchni Issledovatelskaia Rabota or scientific discriminating work (NIR), this system consists of alternating lines of (1) an NTSC-like signal using an amplitude and phase modulated subcarrier and (2) a reference signal having U phase to demodulate the NTSC-like signal. In the linear version, the reference is unmodulated, and in the nonlinear version, the amplitude of the reference signal is modulated with chrominance information.

SUMMARY AND COMPARISONS OF SYSTEMS STANDARDS

History has shown that it is exceedingly difficult to obtain total international agreement on "universal" television broadcasting standards. Even with the first scheduled broadcasting of monochrome television in 1936 in England, the actual telecasting started using two different systems on alternate days from the same transmitter. The Baird system was 250 lines (noninterlaced) with a 50 Hz frame rate while the Electric and Musical Industries (EMI) system was 405 lines (interlaced) with a 25 Hz frame rate.

These efforts were followed in 1939 in the United States by broadcasting a 441 line interlaced system at 60 fields per second (the Radio Manufacturers Association (RMA) system). In 1941, the NTSC initiated the present basic monochrome standards in the U.S. of 525 lines (interlaced) at 60 fields per second, designated as system M by the CCIR. In those early days, the differences in power line frequency were considered as important factors and were largely responsible for the proliferation of different line rates versus field rates as well as the wide variety of video bandwidths. However, the existence and extensive use of monochrome standards over a period of years soon made it a top-priority matter to assume reciprocal compatibility of any developing color system.

The CCIR documents define recommended standards for worldwide color television systems in terms of the three basic color approaches—NTSC, EPAL, and SECAM.[4] The variations—at least 13 of them—are given alphabetical letter designations; some represent major differences while others signify only very minor frequency allocation differences in channel spacings or the differences between the VHF and UHF bands. The key to understanding the CCIR designations lies in recognizing that the letters refer primarily to *local monochrome standards* for line and field rates, video channel bandwidth, and audio carrier relative frequency. Further classification in terms of the particular color system then adds to NTSC, PAL, or SECAM as appropriate. For example, the letter "M" designates a 525-line/60-field, 4.2 MHz bandwidth, 4.5 MHz sound carrier monochrome system. Thus, M(NTSC) describes a color system employing the NTSC technique for introducing the chrominance information within the constraints of the basic monochrome signal values. Likewise, M(PAL) would indicate the same line/field rates and bandwidths but employing the PAL color subcarrier modulation approach.

In another example, the letters "I" and "G" relate to specific 625-line/50-field, 5.0 or 5.5 MHz bandwidth, 5.5 or 6.0 MHz sound carrier monochrome standards. Thus, G(PAL) would describe a 625-line/50-field, 5.5 MHz bandwidth, color system utilizing the PAL color subcarrier modulation approach. The letter "L" refers to a 625-line/50-field, 6.0 MHz bandwidth system to which the SECAM color modulation method has been added (often referred to as SECAM III). System E is an 819-line/50-field, 10 MHz bandwidth, monochrome system. This channel was used in France for early SECAM tests and for system E transmissions.

Some general comparison statements can be made about the underlying monochrome systems and existing color standards:

- There are three different scanning standards: 525-lines/60-fields, 625-lines/50-fields, and 819-lines/50-fields.
- There are six different spacings of video-to-sound carriers, namely 3.5, 4.5, 5.5, 6.0, 6.5, and 11.15 MHz.
- Some systems use FM and others use AM for the sound modulation.
- Some systems use positive polarity (luminance proportional to voltage) modulation of the video carrier while others, such as the U.S. (M)NTSC system, use negative modulation.
- There are differences in the techniques of color sub-

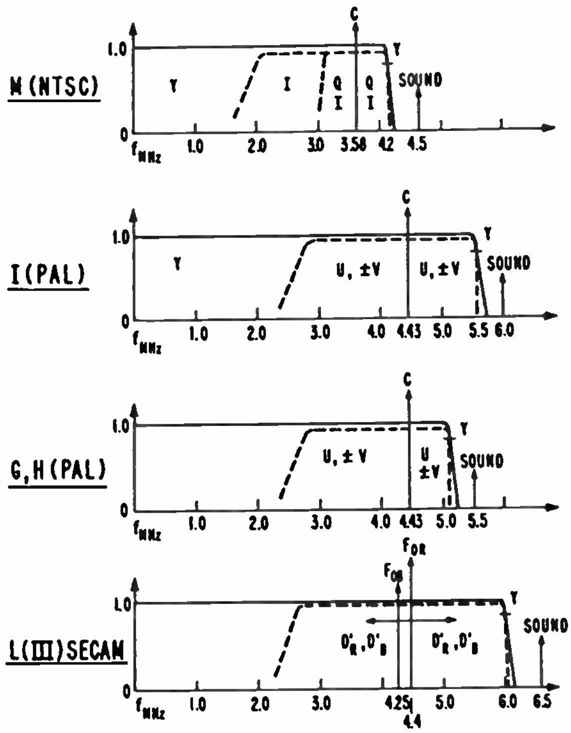


Figure 1.10-13. Bandwidth comparison between NTSC, PAL, and SECAM.

carrier encoding represented by NTSC, PAL, and SECAM, and of course, in each case there are many differences in the details of various pulse widths, timing and tolerance standards.

It is evident that one must refer to the ITU documents for accurate information on the combined monochrome/color standards. Figure 1.10-13 presents a comparison of the relative bandwidths, color subcarrier frequencies, and sound carrier spacing for the major color systems used in the world today.

The signal in the M(NTSC) system occupies the least total channel width, which when the vestigial sideband plus guard bands are included, requires a minimum radio frequency channel spacing of 6 MHz. The L(III) SECAM system signal occupies the greatest channel space with a full 6 MHz luminance bandwidth. Signals from the two versions of PAL lie in between and vary in vestigial sideband width as well as color and luminance bandwidths. NTSC is the only system to incorporate the *I*, *Q* color acuity bandwidth variation. PAL minimizes the color quadrature phase distortion effects by line-to-line averaging, and SECAM avoids this problem by only transmitting the color components sequentially at a line-by-line rate.

Figures 1.10-14–16 summarize, in organization chart form, the CCIR designations for NTSC, PAL, and SECAM basic system identifications and characteristics. In Figure 1.10-14, M(NTSC) identifies the system used in the United States, Canada, Japan, Mexico, the Philippines and several other Central American and Caribbean area countries. The N system may be implemented in color either in the NTSC or the PAL format.[3] Figure 1.10-15 provides a Pritchard summary of the PAL systems. PAL systems in one or another of the 625-line formats are predominately used in Continental Europe, the United Kingdom, some African countries and China. An "M" (525-line) version of PAL has been in use in Brazil.

Figure 1.10-16 summarizes the SECAM III system, which is in use primarily in France and the USSR. The SECAM IV system, as a proposal almost gained favor in 1966 as a universal European approach but was never implemented.[1] The E system, mentioned in connection with early SECAM tests in France, is limited to monochrome broadcasts.

Table 1.10-3 provides a summary of the major color television system general characteristics. Table 1.10-4 characterizes the fundamental features relating to the differences between NTSC, PAL, and SECAM in the critical areas of color encoding techniques. Similarly,

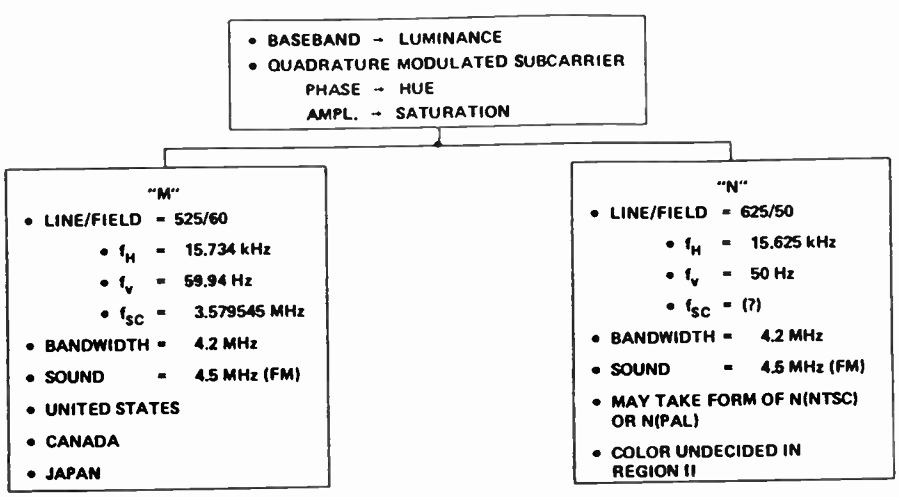


Figure 1.10-14. CCIR designation for NTSC system—summary.

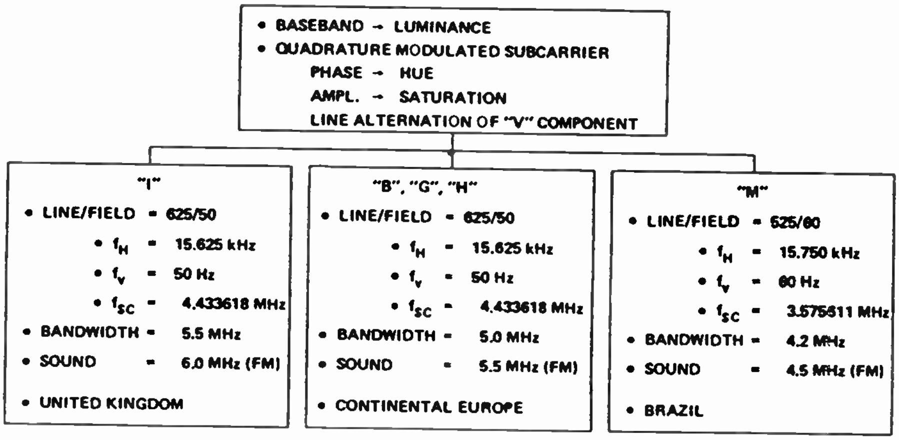


Figure 1.10-15. CCIR designation for PAL system—summary.

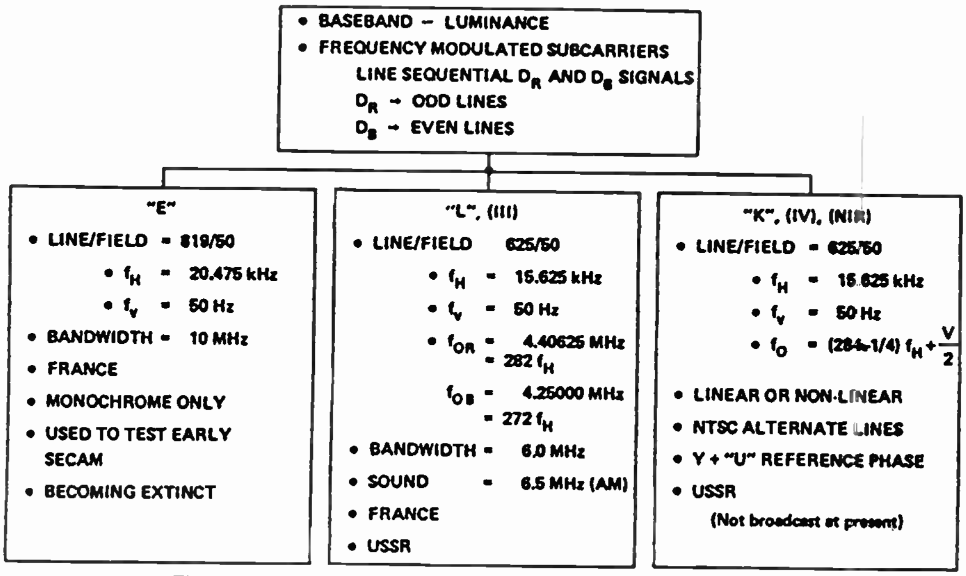


Figure 1.10-16. CCIR designation for SECAM system—summary.

	NTSC	PAL	SECAM
TV SYSTEM	M	G, I	L
FIELD RATE (f_V Hz)	59.94	50	50
TV LINES	525	625	625
LINE RATE (f_H kHz)	15.734	15.625	15.625
LUMA BANDWIDTH (MHz)	4.2	(5.0) (5.5)	6.0
SOUND (MHz)	4.5 (F3)	(5.5) (6.0) (F3)	6.5 (A3)
VERTICAL INTERLACE	2:1	2:1	2:1
GAMMA	2.2	2.8	2.8
WHITE	ILL. "C" (D6500)	D6500	D6500

Table 1.10-3. General system technical summary.

	NTSC	PAL	SECAM
COLOR SUBCARRIER (MHz)	3.579545	4.433618	4.250000 = f_{OB} 4.406500 = f_{OR}
f_{SC} MULTIPLE OF f_H	$\frac{455}{2} f_H$	$\frac{1135}{4} f_H + \frac{f_v}{2}$	272 f_H = f_{OB} 282 f_H = f_{OR}
CHROMA ENCODING	PHASE & AMP. QUAD. MOD.	PHASE & AMP. QUAD. MOD. (LINE ALTERNATION)	FREQUENCY MODULATION (LINE SEQUENTIAL)
COLOR DIFFERENCE SIGNALS	I, Q, (1.3 MHz) (0.6 MHz)	U, ± V (1.3 MHz) (1.3 Mhz)	D_R (f_{OR}) (> 1.0 MHz) D_B (f_{OB}) (> 1.0 MHz)
COLOR BURST PHASE	-(B-Y)	U and ± V	f_{OR} AND f_{OB} 180° PHASE SWITCH EVERY 3rd LINE AND EVERY FIELD
COLOR SWITCH IDENT.	NOT REQUIRED	SWINGING BURST ± 45°	9 LINES OF D_R AND D_B DURING VERTICAL INTERVAL
ADDITIONAL SIGNALS	NONE	"MEANDER" GATE $f_{H/2}$	$f_{H/2}$, $f_{H/4}$, f_v, $f_{v/2}$

Table 1.10-4. Chrominance encoding systems comparison.

		LINE (N)	LINE (N + 1)	LINE (N + 2)	LINE (N + 3)
NTSC:					
	CHROMA:	I, Q	I, Q	I, Q	I, Q
	BURST PHASE:	-(B-Y)	-(B-Y)	-(B-Y)	-(B-Y)
PAL					
	CHROMA:	U,+ V	U,- V	U,+ V	U,- V
	BURST PHASE:	-U + V = + 135°	-U - V = +225°	-U + V = + 135°	-U - V = +225°
SECAM: (FM)					
	CHROMA:	D_R ± 280 kHz	D_B ± 230 kHz	D_R ± 280 kHz	D_B ± 230 kHz
	BURSTS:	(D_R DEVIATION =	$^{+\,350\text{ kHz}}_{-\,500\text{ kHz}}$)	(D_B DEVIATION =	$^{+\,500\text{ kHz}}_{-\,350\text{ kHz}}$)

CHROMA SWITCH IDENT. LINES DURING VERTICAL INTERVAL

LINE #:	7	8	9	10	11	12	13	14	15
	320	321	322	323	324	325	326	327	328
INDENT SIGNALS:	D_R	D_B	D_R	D_B	D_R	D_B	D_R	D_B	D_R

(NOTE: Phase reversed 180° every 3rd line and every field).

Table 1.10-5. Line-to-line chroma signal sequence comparison.

Table 1.10-5 indicates the color encoding line-by-line color sequence operation for the three systems.

The information conveyed in these charts highlights the technical equalities and differences among the systems and attempts to show some kind of order as an aid to understanding the existing worldwide situation. It serves as well to point out the difficulties of entertaining the notion of a "universal" system.[5]

REFERENCES

1. Herbstreit, J. W., and J. Pouliquen, "International Standards for Color Television," *IEEE Spectrum*, IEEE, New York, March 1967.
2. Fink, Donald G. (ed.), *Color Television Standards*, McGraw-Hill Book Co., New York, 1955.
3. Pritchard, D. H., "U.S. Color Television Fundamentals—A Review," *SMPTE Journal*, SMPTE, White Plains, N.Y., Vol. 86, pp. 819-828, November 1977.
4. "CCIR Characteristics of Systems for Monochrome and Colour Television—Recommendations and Reports," Recommendations 470-1 (1974–1978) of the Fourteenth Plenary Assembly of CCIR in Kyoto, Japan, 1978.
5. Roizen, J., "Universal Color Television: An Electronic Fantasia," *IEEE Spectrum*, IEEE, New York, March 1967.

ACKNOWLEDGEMENT

Reprinted with permission of the Society of Motion Picture and Television Engineers. This was originally published in the *SMPTE Journal.*

1.11
DIGITAL TELEVISION TERRESTRIAL TRANSMISSION STANDARDS

COMPILED FROM ATSC DOCUMENTS
INTRODUCTION BY ARTHUR ALLISON
NAB, WASHINGTON, DC
EDITED BY DAVID H. LAYER
NAB, WASHINGTON, DC

INTRODUCTION

The Federal Communications Commission (FCC), with its December 27, 1996 release of the *Fourth Report & Order* in MM docket 87-268, officially established a new standard for television broadcast service in the United States. This new standard is a digital system, based on that established by the broadcast industry through the Advanced Television Systems Committee (ATSC), and promises to fundamentally change not only the technology but the business of TV broadcasting.[1]

This new digital television (DTV) system is described in a series of standards documents adopted by the ATSC. The purpose of this chapter is to provide some background information on the various aspects of DTV technology and to introduce the reader to those standards that define this technology.

Development of the DTV system has its roots in what was originally the desire to increase the resolution of television by developing high definition television (HDTV) systems, which began to gain momentum in the late 1980"s. By 1987, this desire, along with spectrum pressures, led to broadcasters petitioning the FCC to consider the public policy and technical issues related to a significant resolution upgrade of the NTSC system. The FCC began a proceeding by establishing MM Docket 87-268 in July 1987. Later that year the Advisory Committee on Advanced Television Service (ACATS) was formed by the FCC to consider the issues and advise them on how to proceed with a new TV service. The formal objective for the ACATS was:

> *The Committee will advise the Federal Communications Commission on the facts and circumstances regarding advanced television systems for Commission consideration of technical and public policy issues. In the event that the Commission decides that adoption of some form of advanced broadcast television is in the public interest, the Committee would also recommend policies, standards and regulations that would facilitate the orderly and timely introduction of advanced television services in the United States.*

The ACATS established a Planning Subcommittee, a Systems Subcommittee and an Implementation Subcommittee to meet this objective. Each subcommittee then established subgroups to focus investigations into specific areas.

Back in 1987, digital television systems were only gleams in the eyes of those who would invent them. Over the next two years, proposals for delivery of high definition television were offered to the ACATS. A total of 21 systems were proposed with technologies including simulcast, NTSC augmentation, satellite delivery and current NTSC receiver compatibility-but no digital proposals. The candidates were evaluated, with the best proposals withstanding scrutiny and the others falling by the wayside. Then, just before the final cutoff date for submitting proposals in June 1990, General Instrument Corporation submitted an all-digital system.

The digital revolution would change the world forever. In short order, three more all-digital proposals were submitted. The formal testing process began in 1992 with two analog and four digital systems. Although one analog system remained in the competition through the completion of the test process, the digital systems proved far superior. The four digital contenders were found to have different areas of best performance. These four finalist proponents were about to be authorized to implement improvements when another historic milestone was created. The Chairman of ACATS, Richard Wiley, was able to cause the formation of a consortium of the proponents, which became the Grand Alliance on May 24, 1993. The Grand Alliance was to create a single system that would deliver the "best of the best."

In June 1992, the ACATS offered to document the standard for the new television system. This organization had been established in 1983 by the Joint Committee on Inter-Society Coordination (JCIC) to coordinate

---

[1] The Advanced Television Systems Committee (ATSC) is an international cross industry group consisting of over 120 members (companies and organizations) with significant business interest in standardization of various aspects of advanced television systems.

and develop voluntary national standards for advanced television systems.[2]

Testing of the system resulting from the merged designs was completed by 1994. This Grand Alliance system did deliver the "best of the best" of the competing systems. Meanwhile, multiple technical committees were busy writing a standard that described how to provide each element in the HDTV delivery chain. This work was combined into a set of standards, generally referred to as the ATSC Digital Standard, which were voted upon by the ATSC membership in March 1995 and overwhelmingly passed.

During this finalization period, standardization work continued and expanded the scope of the standard to address digital television more generally by adding the capability to encode and transmit standard definition television (SDTV) scanning formats that were compatible with the already standardized HDTV formats. The enhanced standard included formats that exactly matched computer scanning formats as well as those that were directly backward compatible with digitized NTSC video. The updated standard was again overwhelmingly approved by the membership of the ATSC, in September 1995. In November 1995, the ACATS met for the last time and recommended that the FCC adopt the digital TV standard that had been developed and documented by the ATSC, and had been proven practical by the HDTV prototype built by the Grand Alliance.

Meanwhile, the FCC had been requesting input and proposing direction through the rule making process. Each of the last three of six major *Notices of Proposed Rule Making* (NPRM) addresses one of three major areas for rules. One addresses the policy issues, another the technical standard and the last the channel assignments. The key enabling decision was made in December 1996, with the FCC mandating the use of all but one section of the ATSC DTV Standard for the new broadcast service. This was followed in April 1997 with orders on service rules and DTV channel assignments.

The one exception to the FCC mandate of the ATSC Standard is the definition of a subset of the possible scanning formats. By not mandating the entire standard, the particular choices of picture formats made in the ATSC and FCC ACATS processes were made optional. Thus, it will be legal for a broadcaster to send information in any picture format that he chooses to implement within the overall constraints of the transmission technology. However, the broadcast, consumer electronics and other affected industries have agreed on a set of scanning formats; that agreement is embodied in the ATSC DTV Standard which remains a voluntary industry standard.

[2] The members of the JCIC are the National Association of Broadcasters (NAB), the National Cable Television Association (NCTA), The Electronics Industries Association (EIA), the Society of Motion Picture and Television Engineers (SMPTE) and the Institute of Electrical and Electronic Engineers (IEEE).

Receiver manufacturers understand that it is beneficial to all for receivers to have the capability to decode and display all of the picture formats in the ATSC DTV Standard, and are expected to build sets that will do just that. Many standards that the broadcast industry uses are voluntary, that is, no regulatory authority mandates their use. The set of picture formats in the ATSC DTV Standard provide no less a foundation for successful market development than, for example, the industry standards that enable a D1 tape to be recorded and interchanged from machine to machine, or component video signals to be connected among different vendors' equipment. Voluntary standards form the basis of the interoperation of television equipment today and the picture formats standardized by the industry for DTV broadcasting will now join them for the future.

Organization of DTV Standards

The ATSC DTV Standards are organized to define how to get the information from the broadcaster to the receiver, and to select from (or constrain) the options provided by other applicable standards. ATSC's document numbering system uses "A/nn" for the full committee endorsed documents. The ATSC DTV standards are listed in Table 1.11-1.

Although no one document explicitly defines the relationships of the ATSC Standards, the controlling standard is, in effect, Document A/53. This document, along with Document A/52, has the force of regulation since they were both adopted into the FCC rules (in all but one small part). A/53 defines specific requirements, or references other standards with specific constraints, when the entire referenced standard is not included. The referenced standards typically define extensions or alternatives to the MPEG-2 standard to deliver specific functions.

Since A/53 has not been updated since it was adopted, some of the references therein are to document number(s) of work which later became adopted standards, and some relevant documents are not mentioned at all, if they came into existence after A/53 was

Table 1.11-1
ATSC Standards Documents

ATSC Doc. No.	Title
A/52	Digital Audio Compression (AC-3) Standard
A/53	ATSC Digital Television Standard
A/54	Guide to the Use of ATSC Digital Television Standard
A/55	Program Guide for Digital Television (superceded by A/65)
A/56	System Information for Digital Television (superceded by A/65)
A/57	Program/Episode/Version Identification
A/58	Harmonization with DVB SI in the use of the ATSC Digital Television Standard
A/63	Standard for Coding 25/50 Hz Video
A/64	Transmission Measurement and Compliance Standard for Digital Television
A/65	Program and System Information Protocol for Terrestrial Broadcast and Cable

approved. There is an *Informative Guide*, Document A/54, which served as the source for much of this chapter. There are two optional Standards (A/55 and A/56) which have been superceded by Document A/65. Document A/65 is the mandatory standard for the Program and System Information Protocol (PSIP) that defines how to locate television programs and other services. Document A/57 defines a mandatory numbering process for programs. Document A/58 is a *Recommended Practice* (as opposed to a standard), written so as to reduce the possibility that devices and systems designed to the ATSC standard will conflict with those that meet the European set of DTV standards known as DVB. Document A/63 defines 25/50 Hz video formats, for use by those countries that prefer to use that hierarchy of formats. Document A/64 is the *Transmission Measurement and Compliance Document.*

The ATSC standards define how the signal is structured for transmission, while the ISO-1818 (MPEG-2) standards define what the receiver expects as well as the decoding model. While the structure has to be the same from both views of the system, not being aware of the point of view can introduce confusion to one reading both in order to understand the entire system. These standards state the degree of compliance that is expected from implementations, and are typically divided into two sections—*normative* and *informative.*

Those system aspects or specifics contained in the normative section of the standard contain precise technical definitions or requirements. The informative section, on the other hand, is intended to facilitate understanding and/or show one possible way to implement the standard. It is important to note that no requirements can be deduced from the informative section; they only exist in the normative section.

Within the normative section, there are both required and optional specifications. System features that are *required* must be done (system designers *must* implement these system aspects and/or specifics to be considered in compliance with the standard). Things that are *optional* must be done per the standard, if done at all.

There are two basic types of information which exist in an ATSC DTV-standard system—payloads and labels. *Payloads* are audio (A/52), data (no standard as of yet), and video (A/53, Annex A). *Labels* control and describe how to find the payloads. The main label standard is A/53, Annex C, which defines how the payloads are to be combined in the multiplex. The Standards A/53 Annex B, A/57 and A/65 provide the information describing how to find and associate the data streams into programs or services. The RF transmission standards are found in Annex D of A/53. Figure 1.11-1 shows the inter-relationships of these standards from an origination to transmission perspective.

Another perspective which may help in understanding the material contained in this chapter is that the ATSC DTV system, and therefore to a degree the standards which describe it, are organized like a freight train switchyard with a lot of very small cars. Various payloads are defined, divided into standard size containers and assembled into the train with rules about spacing between the containers and their contents.

There are other standards that describe aspects of digital television transmission. One of the more im-

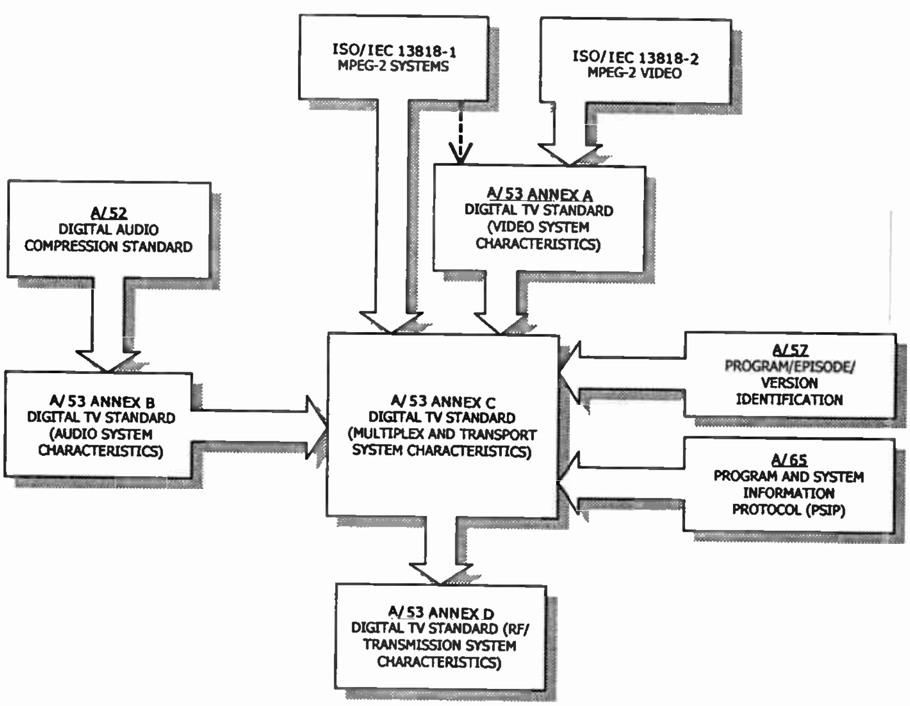


Figure 1.11-1. ATSC DTV Standards—Inter-relationships.

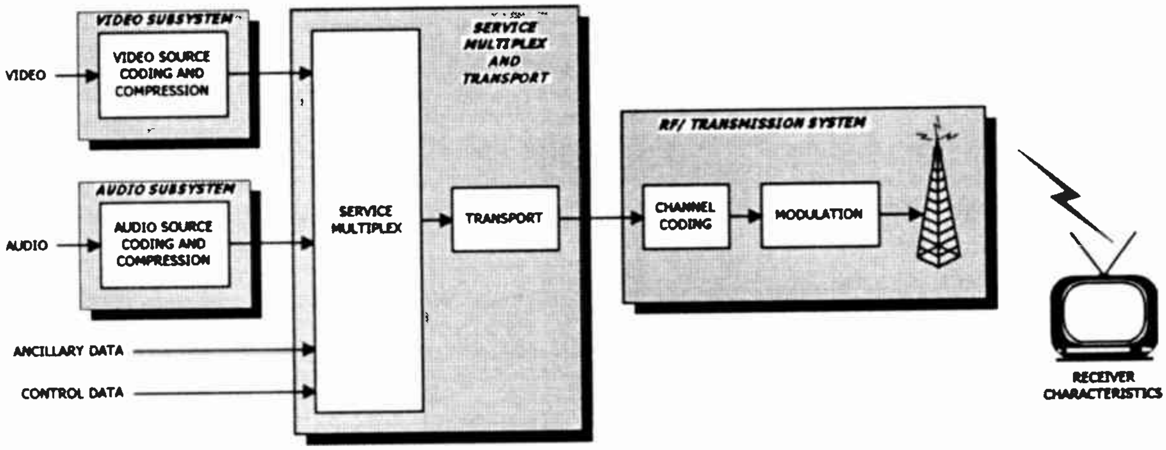


Figure 1.11-2. ITU-R Digital terrestrial television broadcasting model.

portant of these is EIA-708—*Advanced Television Closed Captioning*. EIA-708 describes the contents of the closed captioning container defined by the ATSC standards. This document also references ATSC draft working documents whose contents were incorporated into A/65.

The Society of Cable Telecommunications Engineers (SCTE) adopted standards intended for the cable industry for use in the re-transmission of broadcast DTV signals. These cable standards are different in some aspects with respect to the corresponding broadcast standards. Efforts to reconcile conflicts between differing standards for the same functions are ongoing, and this dynamic situation is not covered in this chapter as a result.

DTV TECHNOLOGY

The following material was excerpted from the various standards and recommended practices documents prepared and adopted by the ATSC, along with the Advanced Television Test Center (ATTC) test reports and FCC determinations as relevant. This material is provided both as an introduction to the technology which forms the basis for DTV as well as for an introduction to the standards themselves. Space limitations prevent complete inclusion of every aspect of these standards. Interested readers are encouraged to consult the actual standards documents for the most complete information, available in original form from the ATSC both in hard-copy and electronic forms.

System Overview

The ATSC DTV Standard describes a system designed to transmit high quality video, audio and ancillary data over a single 6 MHz channel. The system can deliver reliably about 19.4 Mbps of throughput in a 6 MHz terrestrial broadcasting channel. This means that encoding a video source whose resolution can be as high as 5 times that of conventional television (NTSC) resolution requires a bit rate reduction by a factor of 50 or higher. To achieve this bit rate reduction, the system is designed to be efficient in utilizing available channel capacity by exploiting complex video and audio compression technology.

The objective is to maximize the information passed through the data channel by minimizing the amount of data required to represent the video image sequence and its associated audio. This requires representing the video, audio and data sources with as few bits as possible while preserving the level of quality required for the given application.

A basic block diagram representation of the system is shown in Figure 1.11-2, based on a diagram adopted by the International Telecommunication Union, Radiocommunication Sector (ITU-R), Task Group 11/3 (Digital Terrestrial Television Broadcasting). According to this model, the digital television system can be seen to consist of three subsystems.[3]

1. Source coding and compression;
2. Service multiplex and transport; and
3. RF/Transmission.

Source coding and compression refers to the bit rate reduction methods, also known as data compression, appropriate for application to the video, audio and ancillary digital data streams. The term *ancillary data* includes control data, conditional access control data, and data associated with the program audio and video services, such as closed captioning. Ancillary data can also refer to independent program services. The purpose of the coder is to minimize the number of bits needed to represent the audio and video information. The digital television system employs the MPEG-2 video stream syntax for the coding of video and the Digital Audio Compression (AC-3) Standard for the coding of audio.

Service multiplex and transport refers to the means

[3] ITU-R Document TG11/3-2, *Outline of Work for Task Group 11/3, Digital Terrestrial Television Broadcasting*, June 30, 1992.

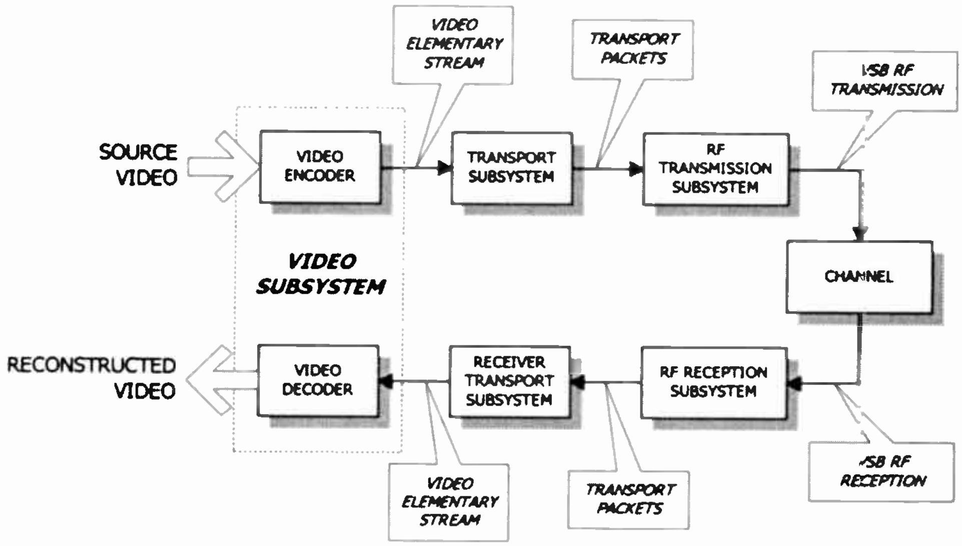


Figure 1.11-3. Video coding in relation to the ATSC DTV system *(Source: ATSC Doc A/54).*

of dividing the digital data stream into packets of information, the means of uniquely identifying each packet or packet type, and the appropriate methods of multiplexing video data stream packets, audio data stream packets, and ancillary data stream packets into a single data stream. The digital television system employs the MPEG-2 transport stream syntax for the packetization and multiplexing of video, audio, and data signals for digital broadcasting systems.[4]

RF/Transmission refers to channel coding and modulation. The channel coder takes the data bit stream and adds additional information that can be used by the receiver to reconstruct the data from the received signal which, due to transmission impairments, may not accurately represent the transmitted signal. The modulation (or physical layer) uses the digital data stream information to modulate the transmitted signal. The modulation subsystem offers two modes: a terrestrial broadcast mode (8 VSB) and a high data rate mode (16 VSB).

VIDEO COMPRESSION

The need for compression of the video information in a digital HDTV system is apparent from the fact that an HDTV studio signal can have 1080 active lines, 1920 samples per line, 10 bits per sample, and 30 frames per second producing a new bit rate of $3 \times 1080 \times 1920 \times 10 \times 30$ or 3 Gbps, but the data capacity of a standard 6 MHz TV channel is approximately 19 Mbps. Consequently, a compression ratio is required. This full data rate is normally converted to 4:2:0 at 8 bits per pixel before compression.

The DTV Standard specifies video compression using a combination of compression techniques, and for reasons of compatibility these compression algorithms have been selected to conform to the specifications of MPEG-2, which is a flexible, internationally accepted collection of compression algorithms. The MPEG-2 specification is organized into a system of profiles and levels, so that applications can ensure interoperability by using equipment and processing that adhere to a common set of coding tools and parameters.[5] The DTV Standard is based on the MPEG-2 Main Profile, which includes three types of video frames (intracoded or I-frames, predicted or P-frames, and bi-directional or B-frames) and an organization of luminance and chrominance samples (designated 4:2:0) within the frame.

The video compression system takes in an analog video source signal and outputs a compressed digital signal that contains information that can be decoded to produce an approximate version of the original image sequence. The goal is for the reconstructed approximation to be imperceptibly different from the original for most viewers, for most images, for most of the time. In order to approach such fidelity, the algorithms are flexible, allowing for frequent adaptive changes in the algorithm depending on scene content, history of the processing, estimates of image complexity and perceptibility of distortions introduced by the compression.

Figure 1.11-3 shows the overall flow of video signals in the ATSC DTV system. A video program source is encoded by a digital television video encoder. The output of the video encoder is a string of bits that represent the video source, and is referred to as a *video elementary stream.* The transport subsystem packetizes the video data into *packetized elementary stream* (PES) packets that are then further packetized into transport packets. The transmission subsystem converts the

[4] Chairman, ITU-R Task Group 11/3, *Report of the Second Meeting of ITU-R Task Group 11/3, Geneva, October 13–19, 1993,* January 5, 1994, p. 40.

[5] For more information about profiles and levels see ISO/IEC 13818-2, Section 8.

Table 1.11-2
Compression Formats (Table 3 of ATSC A/53)

Vertical lines	Pixels	Aspect ratio	Picture rate
1080	1920	16:9	60I, 30P, 24P
720	1280	16:9	60P, 30P, 24P
480	704	16:9	60P, 60I, 30P, 24P
480	704	4:3	60P, 60I, 30P, 24P
480	640	4:3	60P, 60I, 30P, 24P

transport packets into a modulated RF signal for transmission to the receiver. At the receiver, the received signal is demodulated by the receiver transmission subsystem. The receiver transport subsystem converts the received video packets back into a video elementary stream that is decoded by the digital television video decoder. The partitioning shown is conceptual, and practical implementations may differ.

Table 1.11-2 lists the compression formats allowed in the DTV Standard. This table contains the same information as Table 3 in A/53, which has the precise syntax and parameters.

In Table 1.11-2, *vertical lines* refers to the number of active lines in the picture. *Pixels* refers to the number of pixels contained in the active line. *Aspect ratio* refers to the picture aspect ratio *Picture rate* refers to the number of frames or fields per second. In the values for picture rate, "P" refers to progressive scanning; "I" refers to interlaced scanning. Note that each picture rate shown has a dual rate which is 1000/1001 times the rate given in the table (only the base rates are shown in the table for simplicity). It is expected that all frame rates will be at the dual rate as long as NTSC is broadcast due to the complexities of maintaining two timing references in a studio. Document A/63 defines additional video formats for countries which have historically used 50 Hz systems.

Samples are typically obtained using analog-to-digital converter circuits with 8-bit precision. After preprocessing, the various luminance and chrominance samples will typically be represented using 8 bits per sample of each component.

The input video source to the ATSC DTV video compression system is typically in the form of RGB components matrixed into luminance (Y) and chrominance (Cb and Cr) components using a linear transformation (3-by-3 matrix, specified in the standard). The luminance component represents the intensity, or black-and-white picture, while the chrominance components contain color information. The original RGB components are highly correlated with each other. The resulting Y, Cb, and Cr signals have less correlation and are thus easier to code efficiently. The luminance and chrominance components correspond to functioning of the biological vision system; that is, the human visual system responds differently to the luminance and chrominance components.

The coding process may take advantage also of the differences in the ways that humans perceive luminance and chrominance. In the Y, Cb, Cr color space, most of the high frequencies are concentrated in the Y component; the human visual system is less sensitive to high frequencies in the chrominance components than to high frequencies in the luminance component. To exploit these characteristics the chrominance components are low-passed filtered in the ATSC DTV video compression system and sub-sampled by a factor of two along both the horizontal and vertical dimensions, producing chrominance components that are one-fourth the spatial resolution of the luminance component.

Figure 1.11-4 illustrates the spatial relationship between chrominance and luminance samples. For every four luminance samples, there are one each of the Cb and Cr chroma samples. The Cb and Cr chroma samples are located in the same place.

Note that the vertical spatial location of chrominance samples does not correspond to an original sample point, but lies halfway between samples on two successive lines. The 4:2:0 sampling structure thus requires the Cb and Cr samples to be interpolated. For progressively scanned source pictures the processor may simply average the two adjacent (upper and lower) values to compute the sub-sampled values. In the case of interlaced pictures, it can be seen from Figure 1.11-4 that the vertical positions of the chrominance samples in a field are not halfway between the luminance samples of the same field. This is done so that the spatial locations of the chrominance samples in the frame are the same for both interlaced and progressive sources.

The pixels are organized into blocks for the purpose

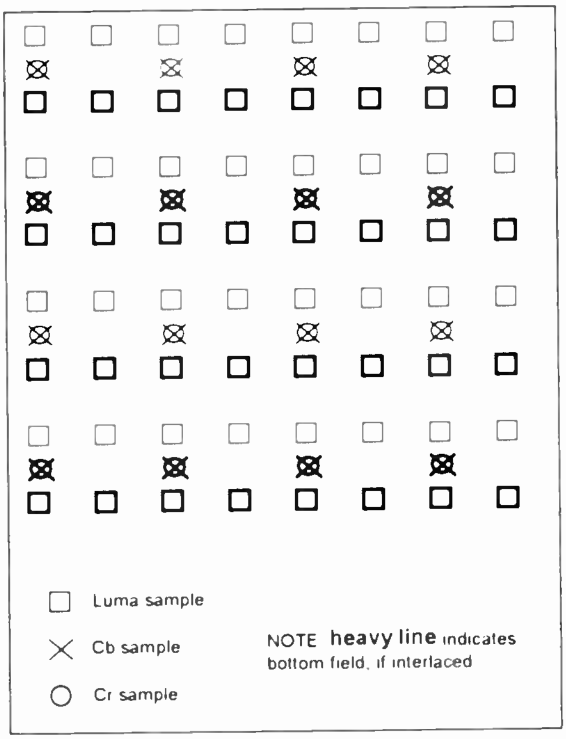


Figure 1.11-4. Placement of luma/chroma samples for 4:2:0 *(Source: ATSC Doc A/54).*

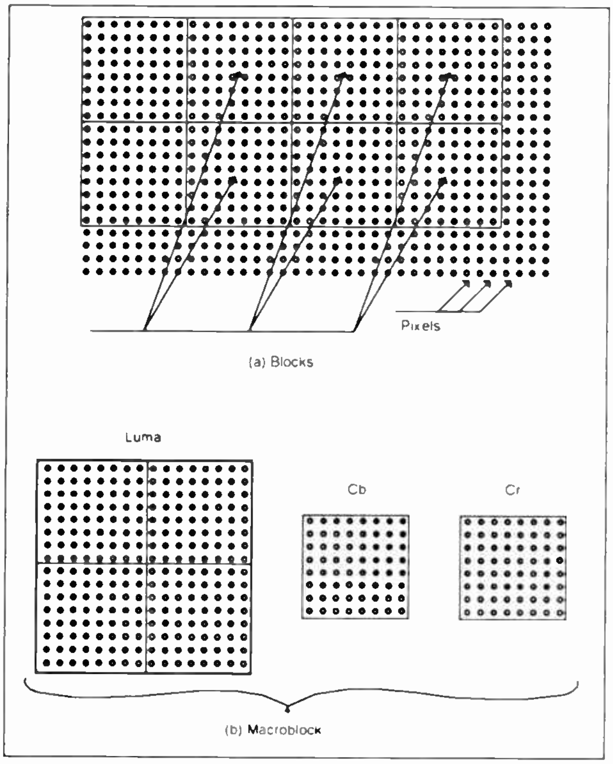


Figure 1.11-5. Blocks and macroblocks *(Source: ATSC Doc A/54).*

of further processing. A block consists of an array of pixel values or an array that is some transform of pixel values. A block for the ATSC DTV system is defined as an array of 8-by-8 values representing either luminance or chrominance information (see Figure 1.11-5a).

Blocks of information are organized into *macroblocks.* A macroblock consists of four blocks of luminance (or a 16 pixel by 16 line region of values) and two chroma (Cb and Cr) blocks. The term macroblock may be used to refer directly to picture element (pel) data or to the transformed and coded representation of pel data. As shown in Figure 1.11-5b, this yields 256 luminance samples and 64 Cb samples and 64 Cr samples (total of 384) per macroblock.

For the 1080 line format (with 1920 samples per line), there are 68 rows of macroblocks (including the row that adds eight dummy lines to create the 1088 lines for coding), with 120 macroblocks per row. For the 720 line format (with 1280 samples per line), there are therefore 45 rows of macroblocks, with 80 macroblocks per row. For the 480 line format with 704 samples per line, there are 30 rows of macroblocks, with 44 macroblocks per row. For the 480 line format with 640 samples per line, there are 30 rows of macroblocks, with 40 macroblocks per row. Horizontal alignment with systems using 720 horizontal samples was in dispute when this chapter was written.

The video coding system requires that the coded picture area has a number of lines that is a multiple of 32 for an interlaced format, and a multiple of 16 for a non-interlaced format. This means that for encoding the 1080 line format, a coder must actually deal with 1088 lines ($1088 = 32 \times 34$). The extra eight lines are in effect dummy lines having no content, and the coder designers will choose dummy data that simplifies the implementation. The extra eight lines are always the last eight lines of the encoded image. Note that the normative standard does not specify top or bottom justified pictures, but the informative states top alignment should be used.

One or more contiguous macroblocks within the same row are grouped together to form *slices.* The order of the macroblocks within a slice is the same as the conventional television *raster* scan, being from left to right. Slices provide a convenient mechanism for limiting the propagation of errors. Since the coded bit stream consists mostly of variable-length codewords, any uncorrected transmission errors will cause a decoder to lose its sense of codeword alignment. Each slice begins with a *slice start* code. Since the MPEG codeword assignment guarantees that no legal combination of codewords can emulate a start code, the slice start code can be used to regain the sense of codeword alignment after an error. When an error occurs in the data stream, the decoder can thus skip to the start of the next slice and resume correct decoding.

The number of slices affects the compression efficiency, partitioning the data stream to have more slices provides for better error recovery but uses bits that could otherwise be used to improve picture quality. The slice is the minimum unit for resynchronization after an error. In the ATSC DTV system, the initial macroblock of every horizontal row of macroblocks is also the beginning of a slice, with possibly several slices across the row.

The primary coding unit of a video sequence is the individual video frame or picture. A video picture consists of the collection of slices that constitute the active picture area. Figure 1.11-6 illustrates a time sequence of video frames consisting of intracoded pictures (I-frames), predictive coded pictures (P-frames) and bi-directionally predictive coded, pictures (B-frames).

Some elements of the compression process exploit only the spatial redundancy within a single picture (frame or field). These processes are called *intraframe coding,* and do not take advantage of the temporal correlation addressed by temporal prediction, which is referred to as *interframe coding.* Frames that do not use any interframe coding are referred to as I-frames (where "I" denotes *intraframe* coded). The ATSC DTV video compression system utilizes both intraframe coding and interframe coding.

The use of periodic I-frames facilitates receiver initializations and channel acquisition (when the receiver is turned on or the receiver channel is changed). The decoder can take advantage of the intraframe coding mode when non-correctable channel errors occur. With motion compensated prediction, an initial frame must be available at the decoder to start the prediction loop. Therefore, a mechanism must be built into the system

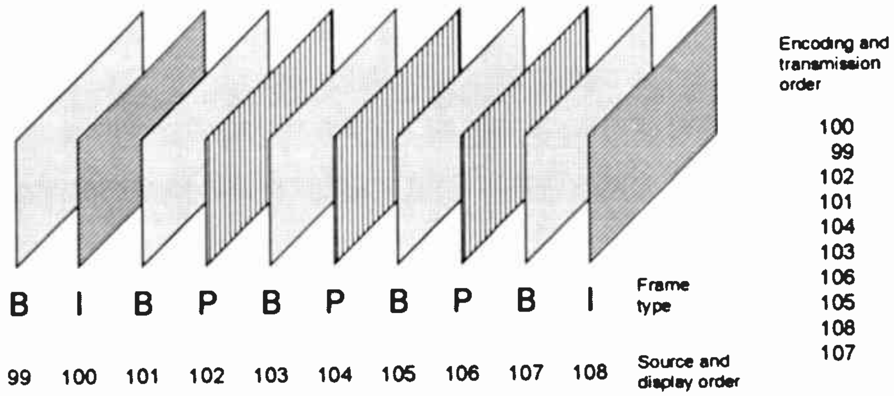


Figure 1.11-6. Illustration of I-frames, P-frames, and B-frames *(Source: ATSC Doc A/54).*

so that if the decoder loses synchronization for any reason, it can rapidly reacquire tracking. The frequency of occurrence of I-pictures may vary and is selected at the encoder. This allows consideration to be given to the need for random access and the location of scene cuts in the video sequence.

P-frames (where "P" denotes *predicted)* are frames where the temporal prediction is in the forward direction only (predictions for the P-frame are formed only from pixels in the most recently decoded I or P-frame). These forward predicted frames allow the exploitation of interframe coding techniques to improve the overall compression efficiency and picture quality. P-frames may include portions that are only intraframe coded. Each macroblock within a P-frame can be either forward-predicted or intraframe coded.

The B-frame (where "B" denotes *bi-directionally* predicted) is a picture type within the coded video sequence that includes prediction from a future frame as well as from a previous frame. The referenced future or previous frames, sometimes called *anchor* frames, are in all cases either I or P-frames.

The basis of the B-frame prediction is that a video frame is correlated both with frames that occur in the past and frames that occur in the future. Consequently, if a future frame is available to the decoder, a superior prediction can be formed, thus saving bits and improving performance. Some of the consequences of using future frames in the prediction are: the B-frame cannot be used for predicting future frames, the transmission order of frames is different from the displayed order of frames and the encoder and decoder must reorder the video frames thereby increasing the total latency. In the example illustrated in Figure 1.11-6, there is one B-frame between each pair of I/P-frames. Each frame is labeled with both its display order and transmission order. The I and P frames are transmitted out of sequence so the video decoder has both anchor frames decoded and available for prediction.

B-frames are used for increasing the compression efficiency and perceived picture quality when encoding latency is not an important factor. The use of B-frames increases coding efficiency for both interlaced and progressively scanned material. The choice of number of bi-directional pictures between any pair of reference (I or P) frames can be determined at the encoder.

Video Encoder Image Prediction and Transform Coding

The DTV compression algorithm depends on creating an estimate of the image being compressed, and subtracting from the image to be compressed the pixel values of the estimate or prediction. If the estimate is good, the subtraction will leave a very small residue to be transmitted; in fact, if the estimate or prediction were perfect, the difference would be zero for all the pixels in the frame of differences, and no new information would need to be sent (that condition can be approached for still images).

If the estimate is not close to zero for some pixels or many pixels, those differences represent information that needs to be transmitted so the decoder can reconstruct a correct image. The kinds of image sequences that cause large prediction differences include severe motion and/or sharp details.

The video coding system uses *motion compensated* prediction as part of the data compression process. Thus macroblocks in the current frame of interest are predicted by macroblock-sized regions in previously transmitted frames. Motion compensation refers to the fact that the locations of the macroblock-sized regions in the reference frame can be offset to account for local motions. The macroblock offsets are known as *motion vectors,* which are *not* specified by the standard. One possible approach might be to perform an exhaustive search to determine the vertical and horizontal offsets that minimize the total difference between the offset region in the reference frame and the macroblock in the frame to be coded.

The estimation of interframe displacement is calculated with half pel precision, in both vertical and horizontal dimensions. That means the displaced macroblock from the previous frame can be displaced by non-integer displacements, and will require interpolation to compute the values of displaced picture elements at locations not in the original array of samples. Estimates for half-pel locations are computed by averages of adjacent sample values.

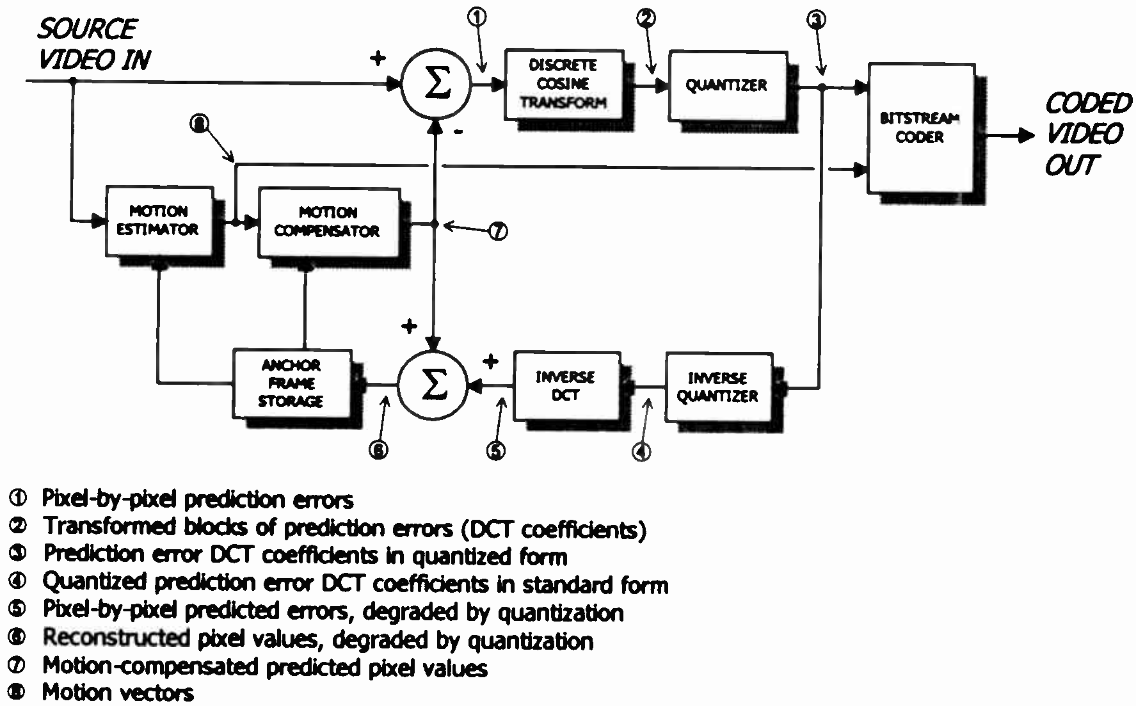


Figure 1.11-7. Encoder prediction loop.

Examination of the encoder prediction loop, shown in Figure 1.11-7, provides insight into the way the different algorithmic elements combine to achieve video compression. This closed feedback loop is the heart of the video compression for the ATSC DTV system.

The prediction loop contains a prediction function that estimates, or predicts, the picture values of the next picture to be encoded in the sequence of successive pictures that constitute the TV program. This prediction is based on previous information that is available within the loop, derived from earlier pictures. The transmission of the predicted compressed information works because the very same information used to make the prediction is available also at the receiving decoder (barring transmission errors, which are expected to be infrequent within the coverage area).

The subtraction of the predicted picture values from the new picture to be coded is at the core of predictive coding. The goal is to do such a good job of predicting the new values that the result of the subtraction function at the beginning of the prediction loop is zero or close to zero for most of the time. The prediction differences are computed separately for the luminance and two chrominance components before further processing. As discussed under I-frames, there are times when prediction is not used, for part of a frame or for an entire frame. Those portions are said to be intraframe coded, while the portions that use the prediction from previous or future pictures are said to be interframe coded.

The prediction differences (sometimes referred to as prediction errors) are grouped into 8-by-8 blocks and a spatial transform is applied to the blocks of difference values. In the intraframe case, the spatial transform is applied to the raw, undifferenced picture data. The luminance and two chrominance components are separately transformed. Since the chrominance data is sub-sampled vertically and horizontally, each 8-by-8 block of chrominance (Cb or Cb) data corresponds to a 16-by-16 macroblock of luminance data, which is not sub-sampled.

The spatial transform used is the *discrete cosine transform* (DCT). The formula for transforming the data is given by:

$$F(u, v) = \frac{1}{4} C(u)C(v)\sum_{x=0}^{7}\sum_{y=0}^{7} f(x, y) \cos\left[\frac{(2x + 1)u\pi}{16}\right] \times \cos\left[\frac{(2y + 1)v\pi}{16}\right]$$

where x and y are pixel indices within an 8-by-8 block, u and v are DCT coefficient indices within an 8-by-8 block, and:

$$C(w) = \frac{1}{\sqrt{2}} \quad \text{for } w = 0$$

$$C(w) = 1 \quad \text{for } w = 1, 2, \ldots, 7$$

Thus an 8-by-8 array of numbers $f(x, y)$ is the input to a mathematical formula, and the output is an 8-

by-8 array of different numbers, $F(u, v)$. The inverse transform[6] is given by:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u, v) \cos\left[\frac{(2x + 1)u\pi}{16}\right] \times \cos\left[\frac{(2y + 1)v\pi}{16}\right]$$

In principle, applying the inverse discrete cosine transform (IDCT) transform to the transformed array would yield exactly the same array as the original. In that sense, transforming the data does not modify the data but merely represents the data in a different form.

The decoder uses the inverse transformation to approximately reconstruct the arrays that were transformed at the encoder, as part of the process of decoding the received compressed data. The approximation in that reconstruction is controlled in advance during the encoding process so as to minimize the effects of coefficient inaccuracies as perceived by the viewer, while simultaneously reducing the quantity of data that needs to be transmitted.

The process of transforming the original data organizes the information in a way that exposes the spatial frequency components of the images or image differences. Using knowledge about the response of the human visual system to different spatial frequencies, the encoder can selectively adjust the precision of transform coefficient representation. The goal is to include as much information about a particular spatial frequency as needed (and possible, given constraints on data transmission), but not to use more precision than is needed, based on visual perception criteria.

Coefficient Quantization and Entropy Coding

The DCT of each 8-by-8 block of pixel values produces an 8-by-8 array of DCT coefficients. The relative precision accorded to each of the 64 DCT coefficients can be selected according to its relative importance in human visual perception. A quantizer matrix represents the relative coefficient precision information, which is an 8-by-8 array of values. Each value in the quantizer matrix represents the coarseness of quantization of the related DCT coefficient.

Two types of quantizer matrices are supported—one which is used for macroblocks which are intraframe coded, and the other for non-intraframe coded macroblocks. The video coding system defines default values for both the intra-quantizer and the *non-intra-quantizer* matrices. Either or both of the quantizer matrices can be overridden at the picture level by transmitting the appropriate arrays of 64 values. Any quantizer matrix overrides stay in effect until the following sequence start code.

The transform coefficients, which represent the bulk of the actual coded video information, are quantized to various degrees of coarseness. As previously indicated the appearance of some portions of the picture will be more affected than others to the loss of precision through coefficient quantization. This phenomenon is exploited by the availability of the quantizer scale factor (used in addition to the coefficient scaling done with the quantizer matrices), allowing the overall level of quantization to vary for each macroblock. Thus entire macroblocks which are deemed to be visually less important can be quantized more coarsely, which results in decreasing the number of bits needed to represent the picture.

For each coefficient other than the DC coefficient of intraframe coded blocks, the quantizer scale factor is multiplied by the corresponding value in the appropriate quantizer matrix to form the quantizer step size. Quantization of the DC coefficients of intra-coded blocks is unaffected by the quantizer scale factor, and is only governed by the (0, 0) element of the intra-quantizer matrix, which is always set to be 8 by ISO/IEC 13818-2.

An important effect of the quantization of transform coefficients is that many coefficients will be rounded to zero after quantization. In fact, a primary method of controlling the encoded data rate is the control of quantization coarseness, since a coarser quantization leads to an increase in the number of zero-value quantized coefficients.

Quantization creates an efficient discrete representation for the data to be transmitted. Codeword assignment takes the quantized values and produces a digital bit stream for transmission. Hypothetically, the quantized values could be simply represented using uniform or fixed-length codewords. Under this approach, every quantized value would be represented with the same number of bits. Greater efficiency, in terms of bit rate, can be achieved by employing entropy coding. Entropy coding attempts to exploit the statistical properties of the signal to be encoded. A signal, whether it contains a pixel value or a transform coefficient, has a certain amount of information, or entropy, based on the probability of the different possible values or events occurring. For example, an event that occurs infrequently conveys much more new information than one that occurs often. By realizing that some events occur more frequently than others, the average bit rate may be reduced.

Huffman coding, which is utilized in the DTV video compression system, is one of the most common entropy coding schemes. In Huffman coding, a codebook is generated which can approach the minimum average description length (in bits) of events, given the probability distribution of all the events. Events that are more likely to occur will be assigned shorter length codewords while those that are less likely to occur will be assigned longer length codewords.

In video compression, most of the transform coefficients are frequently quantized to zero. There may be a few non-zero low frequency coefficients and a sparse scattering of non-zero high frequency coefficients, but

[6] The IDCT is required to conform to *IEEE Standard Specifications for the Implementation of 8×8 Inverse Discrete Cosine Transform*, Std 1180-1990, December 6, 1990.

the great majority of coefficients may have been quantized to zero. To exploit this phenomenon the two-dimensional array of transform coefficients is reformatted and prioritized into a one-dimensional sequence through either a zigzag or alternate scanning process. This results in most of the important non-zero coefficients (in terms of energy and visual perception) being grouped together early in the sequence. They will be followed by long runs of coefficients that are quantized to zero. These zero valued coefficients can be efficiently represented through run-length encoding.

In run-length encoding, the number (run) of consecutive zero coefficients before a non-zero coefficient is encoded, followed by the non-zero coefficient value. The run-length and the coefficient value can be entropy coded, either separately or jointly. The scanning separates most of the zero and the non-zero coefficients into groups, thereby enhancing the efficiency of the run-length encoding process. Also, a special end-of-block (EOB) marker is used to signify when all of the remaining coefficients in the sequence are equal to zero. This approach can be extremely efficient, yielding a significant degree of compression.

As previously indicated above, the array of 64 DCT coefficients is arranged in a one-dimensional vector before run-length/amplitude codeword assignment. Two different one-dimensional arrangements, or *scan types,* are allowed, which are generally referred to as zigzag scan (shown in Figure 1.11-8a) and alternate scan (shown in Figure 1.11-8b). The scan type is specified before coding each picture, and is permitted to vary from picture-to-picture.

Whenever entropy coding is employed, the bit rate produced by the encoder is variable and is a function of the video statistics. Since the bit rate permitted by the transmission system is less than the peak bit rate produced by the variable length coder, a *channel buffer* is necessary at the decoder. The buffering must be carefully designed. Some form of buffer control is necessary to allow efficient allocation of bits to encode the video while ensuring that no overflow or underflow occurs.

The buffer control typically involves a feedback mechanism to the compression algorithm whereby the amplitude resolution (quantization) and/or spatial, temporal and color resolution may be varied in accordance with the instantaneous bit rate requirements. If the bit rate decreases significantly, a finer quantization can be performed to increase it.

As previously indicated the peak bit rate produced by variable length coding will typically fluctuate dramatically and frequently exceed the desired peak channel bandwidth allocation.[7] This situation is handled by the presence of a channel buffer at each decoder for temporary storage of the coded bit stream. The DTV Standard specifies a channel buffer size of 8 Mbits.

A model buffer is defined in the video coding system

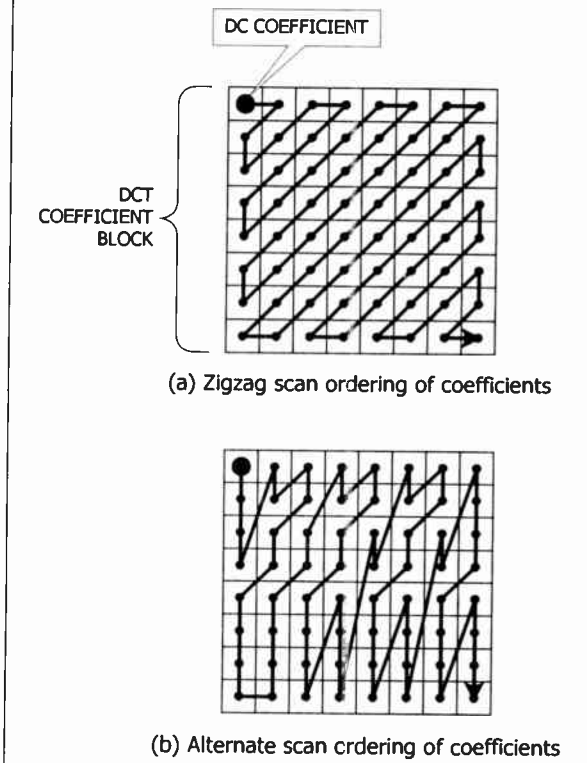


Figure 1.11-8. Scanning of coefficient blocks.

as a reference for manufacturers of both encoders and decoders to ensure interoperability. An encoder will control its production of bits so that the model buffer does not overflow or underflow.[8] In order to avoid overflow or underflow of the model buffer, an encoder may maintain measures of buffer occupancy and scene complexity. When the encoder needs to reduce the number of bits produced, it may do so by increasing the general value of the quantizer scale, which will increase picture degradation. When it is able to produce more bits, it may decrease the quantizer scale thus decreasing picture degradation.

At the decoder the coded coefficients are decoded and an 8-by-8 block of quantized coefficients is reconstructed. Each of these 64 coefficients is *inverse quantized* according to the prevailing quantizer matrix, quantizer scale and frame type. The result of inverse quantization is a block of 64 DCT coefficients.

The decoded and inverse quantized coefficients are organized as 8-by-8 blocks of DCT coefficients and the IDCT is applied to each block. This results in a

[7] Note that the average bit rate, by definition, cannot be permitted to exceed the peak allocated channel bandwidth.

[8] Buffer underflow is actually permitted in the case of low delay bit streams which do not contain B-frames. In such cases underflow might occur due to an unusually difficult picture which requires a particularly large number of bits. The result of underflow is the repeat of one or more pictures, which are presented in lieu of pictures that were skipped in the encoder.

new array of pixel values, or pixel difference values that correspond to the output of the subtraction at the beginning of the prediction loop. If the prediction loop was in the interframe mode, the values will be pixel differences. If the mode was in the intraframe mode, then the inverse transform will produce pixel values directly.

If a portion of the image has not moved, then it is easy to see that a subtraction of the old portion from the new portion of the image will produce zero or nearly zero pixel differences, which is the goal of the prediction. If there has been movement in the portion of the image under consideration, the direct pixel-by-pixel differences will in general not be zero, and might be statistically very large. However, the motion in most natural scenes is organized, and can in most cases be approximately represented locally as a translation. For this reason the video coding system allows for motion compensated prediction, whereby macroblock sized regions in the reference frame may be translated vertically and horizontally with respect to the macroblock being predicted, to compensate for local motion.

The pixel-by-pixel differences between the current macroblock and the motion compensated prediction are transformed by the DCT and quantized using the composition of the *non-intra-quantizer* matrix and the quantizer scale factor. The quantized coefficients are then coded. In the case of I-frames, the entire frame is encoded without reference to any other coded frames. P-frames are referenced to the most recently decoded I or P-frame. B-frames, however, permit the use of two frames as prediction references. One of the reference frames occurs earlier than the coded frame in display order (which can be used for forward prediction), and the other occurs later in display order (which can be used for backward prediction).

For a given macroblock within a B-frame, the encoder has four options-forward prediction, backward prediction, bi-directional prediction, and intraframe coding. When bi-directional prediction is used, the forward and backward predictors are averaged and then subtracted from the target macroblock to form the prediction error. The prediction error is then transformed, quantized and transmitted in the usual manner. Note that both of the frames used as references in coding a B-frame are coded and transmitted prior to the coding of the actual B-frame. This results in the need for frame reordering within the decoder to produce the proper display order. A given picture may be sent by describing the differences between it and one or two previously transmitted pictures. In order for this scheme to work, there must be some way for decoders to become initialized with a valid picture upon tuning into a new channel, or to become re-initialized with a valid picture after experiencing transmission errors.

The process whereby a decoder becomes initialized or re-initialized with valid picture data without reference to previously transmitted picture information is termed *image refresh.* Image refresh is accomplished by the use of intraframe coded macroblocks. There are two general classes of image refresh that can be used either independently or jointly, periodic transmission of I-frames and progressive refresh.

An alternative method for accomplishing image refresh is to encode only a portion of each picture using the intraframe mode. In this case the intraframe coded regions of each picture should be chosen such that, over the course of a reasonable number of frames, all macroblocks are coded intraframe at least once. In addition constraints might be placed on motion vector values in order to avoid possible contamination of refreshed regions through predictions using unrefreshed regions in an uninitialized decoder.

AUDIO COMPRESSION

The audio compression system conforms to the Digital Audio Compression (AC-3) Standard specified in ATSC Document A/52. The audio compression system is considered a constrained subset of the AC-3 Standard. The constraints are specified in Annex B of the DTV Standard.

As illustrated in Figure 1.11-9, the audio subsystem comprises the audio encoding/decoding function and resides between the audio inputs/outputs and the transport subsystem. The audio encoders are responsible for generating the audio elementary streams which are encoded representations of the baseband audio input signals. The flexibility of the transport system allows multiple audio elementary streams to be delivered to the receiver. At the receiver, the transport subsystem is responsible for selecting which audio streams to deliver to the audio subsystem. The audio subsystem is responsible for decoding the audio elementary streams back into baseband audio.

An audio program source is encoded by a digital television audio encoder. The output of the audio encoder is a string of bits that represent the audio source, and is referred to as an *audio elementary stream.* The transport subsystem packetizes the audio data into PES packets that are then further packetized into transport packets. The transmission subsystem converts the transport packets into a modulated RF signal for transmission to the receiver. At the receiver, the received signal is demodulated by the receiver transmission subsystem. The receiver transport subsystem converts the received audio packets back into an audio elementary stream that is decoded by the digital television audio decoder. The partitioning shown is conceptual, and practical implementations may differ. For example, the transport processing may be broken into two blocks; one to perform PES packetization, and the second to perform transport packetization. Or, some of the transport functionality may be included in either the audio coder or the transmission subsystem.

The audio system accepts baseband audio inputs with up to six audio channels per audio program bit stream. The channelization is consistent with ITU-R Recommendation BS-775, *Multi-channel Stereophonic Sound System With and Without Accompanying Picture.* The six audio channels are: Left, Center,

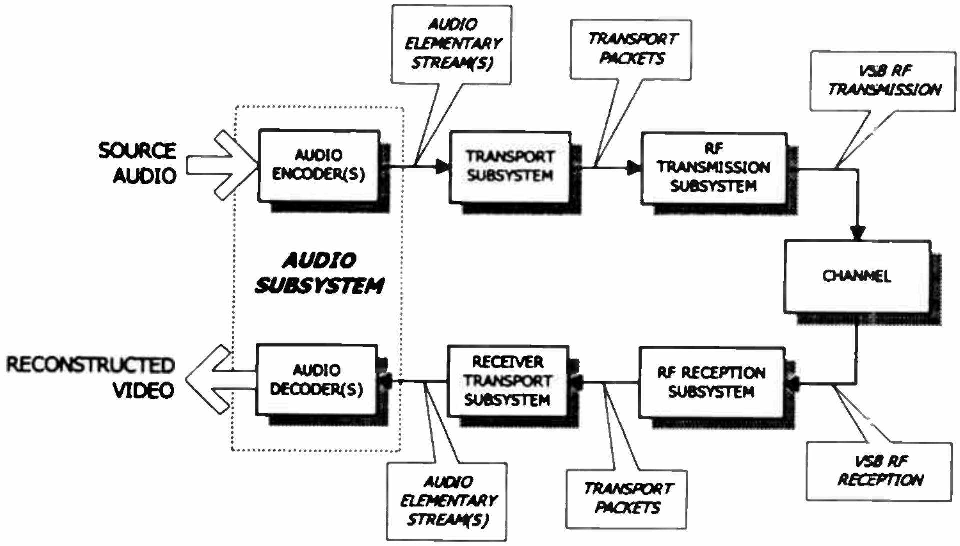


Figure 1.11-9. Audio subsystem within the digital television system.

Right, Left Surround, Right Surround, and Low Frequency Enhancement (LFE). Multiple audio elementary bit streams may be conveyed by the transport system. The bandwidth of the LFE channel is limited to 120 Hz. The bandwidth of the other (main) channels is limited to 20 kHz (this bandwidth difference led to the designation of the 6 channels as 5.1 channels in most documents). Low frequency response may extend to DC, but is more typically limited to approximately 3 Hz (-3 dB) by a DC blocking high-pass filter. Audio coding efficiency (and thus audio quality) is improved by removing DC offset from audio signals before they are encoded.

The system conveys digital audio sampled at a frequency of 48 kHz, locked to the 27 MHz system clock. If analog signal inputs are employed, the A/D converters should sample at 48 kHz. If digital inputs are employed, the input sampling rate shall be 48 kHz, or the audio encoder shall contain sampling rate converters which convert the sampling rate to 48 kHz. The sampling rate at the input to the audio encoder must be locked to the video clock for proper operation of the audio subsystem. In general, input signals should be quantized to at least 16-bit resolution. The audio compression system can convey audio signals with up to 24 bit resolution.

The audio compression system relies upon three basic operations, as shown in Figure 1.11-10. In the first stage, the representation of the audio signal is changed from the time domain to the frequency domain, which is a more efficient domain in which to perform psychoacoustically-based audio compression. The resulting frequency domain coefficients are then encoded. The frequency domain coefficients may be

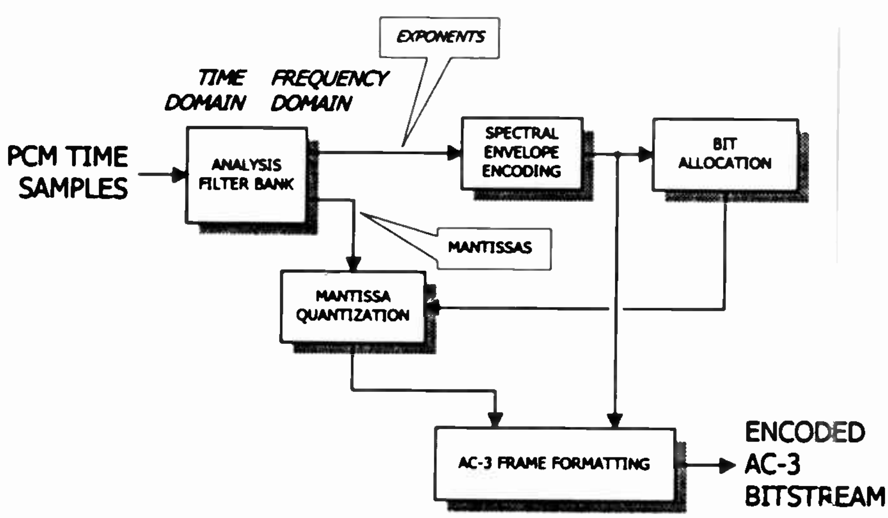


Figure 1.11-10. Overview of audio compression system.

coarsely quantized because the resulting quantizing noise will be at the same frequency as the audio signal, and relatively low signal to noise ratios are acceptable due to the phenomena of psychoacoustic masking. The bit allocation operation determines, based on a psychoacoustic model of human hearing, what actual signal to noise ratio *(SNR)* is acceptable for each individual frequency coefficient.

Finally, the frequency coefficients are coarsely quantized to the necessary precision and formatted into the audio elementary stream. The basic unit of encoded audio is the AC-3 sync frame, which represents 1536 audio samples. Each sync frame of audio is a completely independent encoded entity. The elementary bit stream contains the information necessary to allow the audio decoder to perform the identical (to the encoder) bit allocation. This allows the decoder to unpack and *de-quantize* the elementary bit stream frequency coefficients, resulting in the reconstructed frequency coefficients. The synthesis filterbank is the inverse of the analysis filterbank, and converts the reconstructed frequency coefficients back into a time domain signal.

The process of converting the audio from the time domain to the frequency domain requires that the audio be blocked into overlapping blocks of 512 samples. For every 256 new audio samples, a 512 sample block is formed from the 256 new samples, and the 256 previous samples. Each audio sample is represented in two audio blocks, and thus the number of samples to be processed initially is doubled. The overlapping of blocks is necessary in order to prevent audible blocking artifacts. New audio blocks are formed every 5.33 ms. A group of 6 blocks are coded into one AC-3 sync frame.

Prior to being transformed into the frequency domain, the block of 512 time samples is windowed. The windowing operation involves a vector multiplication of the 512 point block with a 512 point window function. The window function has a value of 1.0 in its center, and tapers down to almost zero at its ends. The shape of the window function is such that the overlap/add processing at the decoder will result in a reconstruction free of blocking artifacts. The window function shape also determines the shape of each individual filterbank filter.

The analysis filterbank is based on the *fast Fourier transform (FFT)*. The particular transformation employed is the oddly stacked *time domain aliasing cancellation (TDAC)* transform. This particular transformation is advantageous because it allows the 100% redundancy that was introduced in the blocking process to be removed. The input to the TDAC transform is 512 windowed time domain points, and the output is 256 frequency domain coefficients.

The frequency coefficients that result from the transformation are converted to a binary floating point notation. The scaling of the transform is such that all values are smaller than 1.0, with 1/65536 resolution. An example value in binary notation (base 2) with 16-bit precision would be:

$$0.0000\ 0000\ 1010\ 1100_2$$

The number of leading zeroes in the coefficient, 8 in this example, becomes the raw exponent. The value is left shifted by the exponent, and the value to the right of the decimal point (1010 1100) becomes the normalized *mantissa* to be coarsely quantized. The exponents and the coarsely quantized mantissas are encoded into the bit stream.

Some processing is applied to the raw exponents in order to reduce the amount of data required to encode them. First, the raw exponents of the six blocks to be included in a single AC-3 sync frame are examined for block-to-block differences. If the differences are small, a single exponent set is generated which is useable by all six blocks, thus reducing the amount of data to be encoded by a factor of six. If the exponents undergo significant changes within the frame, then exponent sets are formed over blocks where the changes are not significant. Due to the frequency response of the individual filters in the analysis filter bank, exponents for adjacent frequencies rarely differ by more than ±2. To take advantage of this fact, exponents are encoded differentially in frequency. The first exponent is encoded as an absolute, and the difference between the current exponent and the following exponent is then encoded. This reduces the exponent data rate by a factor of two. Finally, where the spectrum is relatively flat, or an exponent set only covers 1-2 blocks, differential exponents may be shared across two or four frequency coefficients, for an additional savings of a factor of two or four.

The final coding efficiency for exponents is typically 0.39 bits/exponent (or equivalently, 0.39 bits/sample since there is an exponent for each audio sample). Exponents are only coded up to the frequency needed for the perception of full frequency response. Typically, the highest audio frequency component in the signal that is audible is at a frequency lower than 20 kHz. In the case that signal components above 15 kHz are inaudible, only the first 75% of the exponent values are encoded, reducing the exponent data rate to <0.3 bits/sample.

The exponent processing changes the exponent values from their original values. The encoder generates a local representation of the exponents which is identical to the decoded representation that will be used by the decoder. The decoded representation is then used to shift the original frequency coefficients to generate the normalized mantissas that are quantized.

The precision of the frequency coefficients produced by the analysis filterbank depends on the word length of the input pulse coded modulation (PCM) audio samples and the precision of the transform computation. Typically this precision is on the order of 16-18 bits, but may be as high as 24 bits. Each normalized mantissa is quantized to a precision between 0 and 16 bits. The goal of audio compression is to maximize the audio quality at a given bit rate. This requires an optimum (or near optimum) allocation of the available bits to the individual mantissas.

The number of bits allocated to each individual mantissa value is determined by the bit allocation routine. The identical core routine is run in both the encoder and the decoder, so that each generates the identical bit allocation.

Compressed Audio Bit Stream

The audio bit stream consists of a repetition of audio frames which are referred to as *AC-3 sync frames*. Each AC-3 sync frame is a self contained entity consisting of synchronization information (SI), bit stream information (BSI), 32 ms of encoded audio, and a CRC error check code. Every sync frame is the same size (number of bits) and contains six encoded audio blocks. Within SI is a 16-bit sync word, an indication of audio sample rate (48 kHz for the digital television system) and an indication of the size of the audio frame (which indicates bit rate).

The ideal place to splice encoded audio bit streams is at the boundary of a sync frame. If a bit stream splice is performed at the sync frame boundary, the audio decoding will proceed without interruption. If a bit stream splice is performed randomly, there will be an audio interruption. The frame which is incomplete will not pass the decoder''s error detection test and this will cause the decoder to mute. The decoder will not find sync in its proper place in the next frame, and will enter a sync search mode. Once the sync code of the new bit stream is found, synchronization will be achieved, and audio reproduction may begin once again. The outage will be on the order of two frames, or about 64 ms. Due to the windowing process of the filterbank, when the audio goes to mute there will be a gentle fade down over a period of 2.6 ms. When the audio is recovered, it will fade up over a period of 2.6 ms. Except for the approximately 64 ms of time during which the audio is muted, the effect of a random splice of an AC-3 elementary stream is theoretically relatively benign.

Loudness and Dynamic Range

The digital television system was designed to provide uniform subjective loudness for all audio programs. Audio levels fluctuate between broadcast channels (observed when channel hopping), or between program segments on a particular channel (commercials much louder than the entertainment) in the NTSC system, which consumers find to be very annoying. One element that is found in most audio programming is the human voice. Achieving an approximate level match for dialogue (spoken in a normal voice, not shouting or whispering) amongst all audio programming was a design goal. The AC-3 audio system provides syntactical elements that are expected to make this goal achievable.

Since the digital audio coding system can provide more than 100 dB of dynamic range, there is no technical reason for dialogue to be encoded anywhere near 100% as is commonly done in NTSC television. However, there is no assurance that all program channels, or all programs or program segments on a given channel, will have dialogue encoded at the same (or even similar) level. Lacking a uniform coding level for dialogue (which would imply a uniform headroom available for all programs) there would be inevitable audio level fluctuations between program channels and even between program segments.

Encoded AC-3 elementary bit streams are tagged with an indication of the subjective level at which dialogue has been encoded. Different audio programs may be encoded with differing amounts of headroom above the level of dialogue in order to allow for dynamic music and sound effects. The digital television receiver (and all AC-3 decoders) are able to use this information on dialog level to adjust the reproduced level of audio programs, so that different received programs have their spoken dialogue reproduced at a uniform level. Some receiver designs may even offer the listener an audio volume control calibrated in absolute sound pressure level. The listener could dial up the desired sound pressure level (SPL) for dialogue, and the receiver would scale the level of every decoded audio program so that the dialogue is always reproduced at the desired level.

It is common practice for high quality programming to be produced with wide dynamic range audio, suitable for the highest quality audio reproduction environment. Broadcasters, serving a wide audience, typically process audio in order to reduce its dynamic range. The processed audio is more suitable for the majority of the audience that does not have an audio reproduction environment that matches that of the original audio production studio. In the case of NTSC, all viewers receive the same audio with the same dynamic range, and it is impossible for any viewer to enjoy the original wide dynamic range audio production.

The DTV audio coding system provides an embedded dynamic range control system that allows a common encoded bit stream to deliver programming with a dynamic range appropriate for each individual listener. A dynamic range control value is provided in each audio block (every 5 ms). The audio decoder uses these values in order to alter the level of the reproduced audio for each audio block. Level variations of up to ±24 dB may be indicated. The dynamic range values are generated in order to provide a subjectively pleasing but restricted dynamic range.

Some receivers may offer the viewer the option to scale the dynamic range values in order to reduce the effect of the dynamic range compression that was introduced by the broadcaster. Thus the viewer may be given independent control of the amount of compression applied to loud and quiet sounds. Therefore, while the broadcaster may introduce dynamic range compression to suit the needs of most of the audience, individual listeners may have the option to choose to enjoy the audio program with more or all of its original dynamic range intact.

Service Types

An AC-3 elementary stream contains the encoded representation of a single audio service. Multiple elementary streams provide multiple audio services. Each elementary stream is conveyed by the transport multiplex with a unique *packet identifier* (PID). There are a number of audio service types that may be (individually) coded into each elementary stream. Each elementary stream is tagged as to its service type. There are two types of main service and six types of associated service. Each associated service may be tagged (in the AC-3 audio descriptor in the transport PSI data) as being associated with one or more main audio services. Each AC-3 elementary stream may also be tagged with a language code.

Associated services may contain complete program mixes, or may contain only a single program element. Associated services that are complete mixes may be decoded and used in its present state. Associated services that contain only a single program element are intended to be combined with the program elements from a main audio service.

The service types available are defined in the Digital Audio Compression (AC-3) Standard and in Annex B of the DTV Standard. This information is reproduced in Table 1.11-3 and the following paragraphs briefly describe the meaning of these service types:

Complete main audio service (CM) This is the normal mode of operation. All elements of a complete audio program are present. The audio program may be any number of channels from 1 to 5.1.

Main audio service, music and effects (ME) All elements of an audio program are present except for dialogue. This audio program may contain from 1 to 5.1 channels. Dialogue may be provided by a D associated service (that may be simultaneously decoded and added to form a complete program).

Associated service: visually impaired (VI) This is typically a single-channel service, intended to convey a narrative description of the picture content for use by the visually impaired, and intended to be decoded along with the main audio service. The VI service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

Associated service: hearing impaired (HI) This is typically a single-channel service, intended to convey dialogue which has been processed for increased intelligibility for the hearing impaired, and intended to be decoded along with the main audio service. The HI service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

Associated service: dialogue (D) This service conveys dialogue intended to be mixed into a main audio service (ME) which does not contain dialogue.

Associated service: commentary (C) This service typically conveys a single-channel of commentary intended to be optionally decoded along with the main audio service. This commentary channel differs from a dialogue service, in that it contains optional instead of necessary program content. The C service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

Associated service: emergency message (E) This is a single-channel service, which is given priority in reproduction. If this service type appears in the transport multiplex, it is routed to the audio decoder. If the audio decoder receives this service type, it will decode and reproduce the E channel while muting the main service - this is a mandatory requirement.

Associated service: voice-over (VO) This is a single-channel service intended to be decoded and added into the center loudspeaker channel.

Each audio bit stream may be in any language. In order to provide audio services in multiple languages a number of main audio services may be provided, each in a different language. This is the (artistically) preferred method, because it allows unrestricted placement of dialogue along with the dialogue reverberation. The disadvantage is as much as 384 kbps is needed to provide a full 5.1 channel service for each language. One way to reduce the required bit rate is to reduce the number of audio channels provided for languages with a limited audience. For instance, alternate language versions could be provided in 2 channel stereo with a bit rate of 128 kbps. Or, a mono version can be supplied at a bit-rate of approximately 64-96 kbps.

Formation of a complete audio program requires that the appropriate language D service be simultaneously decoded and mixed into the ME service. This method allows a large number of languages to be efficiently provided, but at the expense of artistic limitations. The single channel of dialogue would be mixed into the center reproduction channel and could not be panned. Also, reverberation would be confined to the center channel, which is not optimum. Nevertheless, for some types of programming (sports, etc.) this method is very attractive due to the savings in bit rate it offers. Another way to offer service in multiple languages is to provide a main multi-channel audio service (ME) which does not contain dialogue. Multiple single channel dialogue associated services (D) can then be provided, each at a bit-rate of approximately 64-96 kbps.

Some receivers may not have the capability to simultaneously decode two audio services.

Table 1.11-3

Table of Service Types

Bsmo d	Type of service
000 (0)	Main audio service: complete main (CM)
001 (1)	Main audio service: music and effects (ME)
010 (2)	Associated service: visually impaired (VI)
011 (3)	Associated service: hearing impaired (HI)
100 (4)	Associated service: dialogue (D)
101 (5)	Associated service: commentary (C)
110 (6)	Associated service: emergency (E)
111 (7)	Associated service: voice-over (VO)

ANCILLARY DATA SERVICES

The digital television system affords the opportunity to augment the basic television video and audio service with ancillary digital data services. The flexibility of the MPEG-2 transport layer employed in the system allows new digital services to be easily introduced at any time in a completely backward compatible manner. Two basic ancillary services were specified as a part of the DTV Standard from the outset. These services were to provide for the transmission of Closed Captioning (mislabeled as Program Subtitles in A/54), ancillary services, and program selection information.

Program Subtitles are completely analogous to the Closed Caption service transmitted on line 21 of the analog NTSC signal. Emergency Messages are mixed into the baseband video signal in the case of NTSC signals. The ancillary services envisioned are a mix of associated audio services and textual services.

The original standards adopted in 1995 provided multiple methods for locating a program. In 1997, a new structured method for tuning to programs was created and documented as an ATSC standard. This mandatory standard provides a way for the broadcaster to announce and the receiver to select a program. It contains time and control information to facilitate navigation and as such it allows suitably equipped receivers to build an interactive on-screen grid of program information. The information about the planned programs is carried in three major tables-a master guide table, a virtual channel table and an event information table. In addition, tables to provide system time, programs ratings and extended text messages about events are also defined. More information about this standard is in the Program and System Information Protocol (PSIP) section of this chapter and in Document A/65, which was approved by the ATSC on December 22, 1997.

When new data services are defined, the rules in Section 5.6.2 of Annex C of the DTV Standard specify that the ancillary services be specified in terms of a verification model by defining the characteristics of the transmitted syntax and an idealized decoder.

In MPEG-2 parlance, these ancillary services are referred to as *private data.* Private data provides a means to add new ancillary services to the basic digital television service specified in this Standard. Private data is supported in two bit stream locations:

1. Private data can be transmitted within the adaptation header of transport packets (Sections 2.4.3.4 and 2.4.3.5 of ISO/IEC 13818-1);
2. Private data can be transmitted as a separate transport stream with its own PID.

In either case, it is necessary that the standards that specify the characteristics of such data be consistent with A/53.

This mechanism is being used by ATSC to define how to provide both one-way and two-way interactive data services.

Transport Mechanism

The precise transport system standard for DTV is based on MPEG-2 and is given in ISO/IEC 13818-1 as constrained and extended for the DTV Standard. The constraints and extensions to the MPEG-2 standard that apply are given in Annex C of the DTV Standard.

The transport system employs the fixed length transport stream packetization approach defined by MPEG. This approach to the transport layer is well suited to the needs of terrestrial broadcast and cable television transmission of digital television. The use of moderately long, fixed length packets matches well with the needs and techniques for error protection in both terrestrial broadcast and cable television distribution environments. At the same time it provides great flexibility to accommodate the initial needs of the service to multiplex video, audio and data while providing a well-defined path to add additional services in the future in a fully backward compatible manner. By basing the transport layer on MPEG-2, maximum interoperability with other media and standards is maintained.

While the transport format conforms to the MPEG-2 Systems Standard, it will not exercise all the capabilities defined in the MPEG-2 Standard. Therefore, a digital television decoder need not be fully MPEG-2 Systems compliant, in that it will not need to decode any arbitrary MPEG-2 Systems bit stream. However, all MPEG-2 decoders should be able to decode the digital television bit stream syntax at the transport system level.

Figure 1.11-11 illustrates the organization of a digital television transmitter-receiver pair and the location of the transport subsystem in the overall system. The transport resides between the application (audio or video) encoding/decoding function and the transmission subsystem. At its lowest layer, the encoder transport subsystem is responsible for formatting the encoded bits and multiplexing the different components of the program for transmission. At the receiver, it is responsible for recovering the bit streams for the individual application decoders and for the corresponding error signaling. (At a higher layer, multiplexing and demultiplexing of multiple programs within a single bit stream can be achieved with an additional system level multiplexing or demultiplexing stage before the modem in the transmitter and after the modem in the receiver.) The transport subsystem also incorporates other higher level functionality related to identification of applications and, as illustrated, synchronization of the receiver.

As described earlier, the data transport mechanism is based on the use of fixed length packets that are identified by headers. Each header identifies a particular application bit stream (also called an elementary bit stream) which forms the payload of the packet. Applications supported include video, audio, data, program and system control information, etc. The elementary bit streams for video and audio are wrapped in a variable length packet structure called the packetized elementary stream (PES) before transport processing. The PES layer provides functionality for identification,

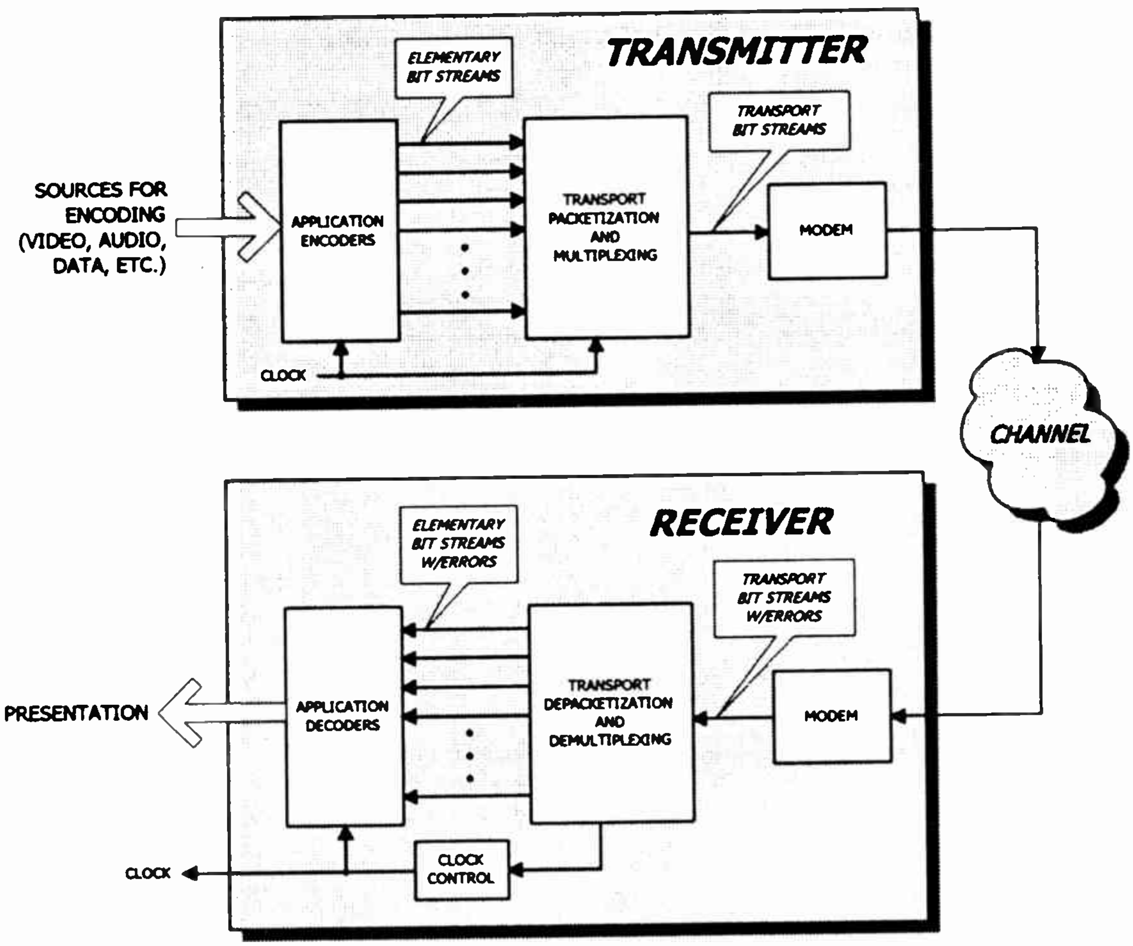


Figure 1.11-11. Sample organization of functionality in a transmitter-receiver pair for a single digital television program.

and synchronization of decoding and presentation of the individual application.

Moving up one level in the description of the general organization of the bit streams, elementary bit streams sharing a common time base are multiplexed, along with a control data stream, into *programs*. Note that a program in the digital television system is somewhat analogous to a channel in the NTSC system in that it contains all of the video, audio, and other information required to make up a complete television program.[9] These programs and an overall system control data stream are then asynchronously multiplexed to form a multiplexed *system*.

At this level, the transport is flexible in two aspects:

- It permits programs to be defined as any combination of elementary bit streams; specifically, the same elementary bit stream can be present in more than one program (two different video bit streams with the same audio bit stream); a program can be formed by combining a basic elementary bit stream and a supplementary elementary bit stream (bit streams for scaleable decoders); programs can be tailored for specific needs (regional selection of language for broadcast of secondary audio); etc;
- Flexibility at the systems layer allows different programs to be multiplexed into the system as desired, and allows the system to be reconfigured when required. The procedure for extraction of separate programs from within a system is also well defined. A/65 defines both required methods as well as optional methods.

The transport system provides other features that are useful for both normal decoder operation and for the special features required in broadcast and cable applications. These include:

- Decoder synchronization
- Conditional access
- Local program insertion

The transport bit stream definition directly addresses issues relating to the storage and playback of programs. Although this is not directly related to the transmission of digital television programs, it is a fundamental re-

[9] If one ignores the data which can reside in the NTSC analog multiplex.

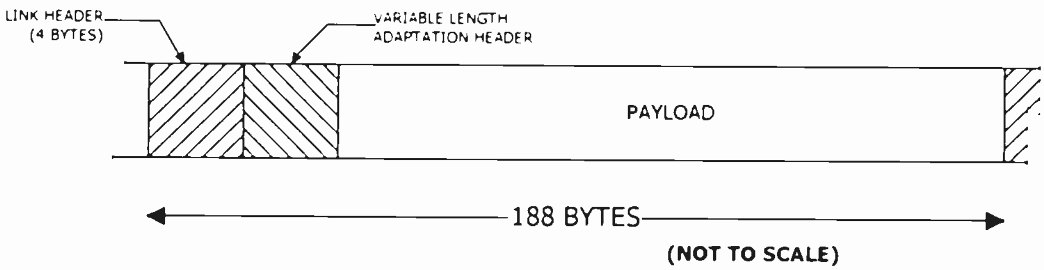


Figure 1.11-12. Transport packet format.

quirement for creating programs in advance, storing them and playing them back at the desired time. The programs are stored in the same format in which they are transmitted, as transport bit streams. The bit stream format also contains the hooks needed to support the design of consumer digital products based on recording and playback of these bit streams, including the use of the trick modes that one is familiar with for current analog VCRs (but no standard methods are defined). It should be noted that the issues related to storage and playback of digitally compressed video bit streams are quite different from those that apply to analog systems such as NTSC.

The transport bit stream consists of fixed length packets with a fixed and a variable component to the header field as illustrated in Figure 1.11-12.

Each packet consists of 188 bytes and is constructed in accordance with the MPEG-2 transport syntax and semantics. The choice of this packet size was motivated by a few factors. The packets need to be large enough so that the overhead due to the transport headers does not become a significant portion of the total data carried. However, they should not be so large that the probability of packet error becomes significant under standard operating conditions (due to inefficient error correction). It is also desirable to have packet lengths consistent with the block sizes of typical, block oriented, error correction methods, so that packets may be synchronized to error correction blocks, and the physical layer of the system can aid the packet level synchronization process in the decoder. Another reason for the particular packet length selection is interoperability with the ATM format. It has been shown that one can transmit a single MPEG-2 transport packet in four ATM cells.

The packet headers identify the contents of each packet. The packet header structure is layered and may be described as a combination of a fixed length link layer and a variable length adaptation layer. Each layer serves a different function similar to the link and transport layer functions in the OSI layered model of a communications system. In the digital television system, this link and adaptation level functionality is used directly for the terrestrial broadcast link on which the MPEG-2 transport bit stream is transmitted. However, in a different communications system (ATM), the MPEG-2 headers would not play a role in implementing a protocol layer in the overall transmission system. The MPEG-2 headers would be carried as part of the payload in such a case and would continue to serve as identifiers for the contents of the data stream.

Random entry into the application bit streams such as video and audio is necessary to support functions such as program acquisition and program switching. Random entry into an application is possible only if the coding for the elementary bit stream for the application supports this functionality directly. For example, the video bit stream supports random entry through the concept of intraframes (or I-frames) that are coded without any prediction, and can therefore be decoded without any prior information. The beginning of the video sequence header information preceding data for an I-frame could serve as a random entry point into a video elementary bit stream. In general, random entry points should also coincide with the start of PES packets where they are used (for video and audio).

The support for random entry at the transport layer comes from a flag in the adaptation header of the packet that indicates whether the packet contains a random access point for the elementary bit stream. In addition, the data payload of packets that are random access points starts with the data that forms the random access point of entry into the elementary bit stream itself. This approach allows packets to be discarded directly at the transport layer when switching channels and searching for a resynchronization point in the transport bit stream, and also simplifies the search for the random access point in the elementary bit stream once transport level resynchronization is achieved.

The transport system supports insertion of local programs and commercials by use of flags and features dedicated to this purpose in the transport packet Adaptation Header. This important issue was discussed at great length within MPEG. The syntax allows local program insertion to be supported and its performance to improve as techniques and equipment are developed around these syntax tools. The use of these syntax elements will need to be within imposed constraints to ensure proper operation of the video decoders. There may also need to be constraints on some current common broadcast practices, imposed not by the transport, but rather by virtue of the compressed digital data format. Significant effort by industry to develop improved switching techniques, to provide broadcasters with at least the same functionality as NTSC switching, are currently underway.

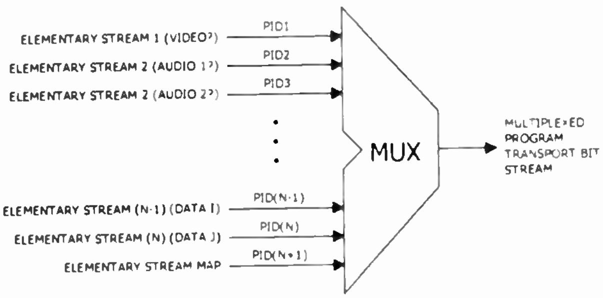


Figure 1.11-13. Illustration of the multiplex function to form a program transport stream.

Multiplexing

A program transport stream is formed by multiplexing individual elementary bit streams (with or without PES packetization) that share a common time base.[10] As the elementary streams are multiplexed, they are formed into transport packets and a control bit stream that describes the program (also formed into transport packets) is added. The organization of this multiplex function is illustrated in Figure 1.11-13. The elementary bit streams, and the control bit stream (also called the elementary stream map in Figure 1.11-13), are identified by their unique PIDs in the link header field. The control bit stream contains the information that describes the elementary stream map, including information about the PIDs of the transport streams that make up the program, the identification of the applications (audio, video, etc.) that are being transmitted on these bit streams, the relationship between these bit streams, etc.

The transport syntax allows a program to be comprised of a large number of elementary bit streams, with no restriction on the types of applications required within a program. For example, a program transport stream does not need to contain a single video or audio bit stream; it could be a data program. On the other hand, a program transport stream could contain multiple related video and audio bit streams, as long as they share a common time base. The data applications that can be carried are flexible; the only constraint is that there should be an appropriate ID assignment to enable recognition of the application corresponding to the bit stream by a decoder.

Note that, for the different elementary bit streams that make up a program, the link level functions are carried out independently without program level coordination. This includes functions such as PID manipulation, bit stream filtering, scrambling and descrambling, definition of random entry packets, etc. The coordination between the elements of a program is primarily controlled at the presentation (display) stage based on the use of the common time base. This common time base is imposed by having all elementary bit streams in a program derive timing information from a single clock, and then by transmitting this timing information via the program clock reference (PCR) on one of the elementary bit streams that constitute the program. The data for timing of presentation is present in the elementary bit stream for each individual application.

The system multiplex is created by assembling different program transport streams in a proscribed manner. In addition to the transport bit streams (with the corresponding PIDs) that define the individual programs, a system level control bit stream with PID = 0 is also defined and included in the multiplex. Although this bit stream carries the information that maps program identities to their program transport streams, the PSIP standard enables determination of the contents of the multiplex before the channel change is attempted, facilitating faster program acquisition (what we used to know as channel change). Again, in this context, a program corresponds to what has traditionally been called a channel (PBS, C-SPAN, etc.).

The system layer of multiplexing is illustrated in Figure 1.11-14. Note that during the process of system level multiplexing, there is the possibility of PIDs on different program streams being identical at the input. This poses a problem because PIDs for different bit streams need to be unique. A solution to this problem lies at the multiplexing stage, where some of the PIDs are modified just before the multiplex operation. Hardware implementation of the PID reassignment function in real time is helped by the fact that this process is synchronous at the packet clock rate. The other approach, of course, is to make sure up front that the PIDs being used in the programs that make up the system are unique. This is not always possible with stored bit streams.

Note that the architecture of the bit stream is scalable. Multiple system level bit streams can be multiplexed together on a higher bandwidth channel by extracting the appropriate information from each system multiplexed bit stream and reconstructing a new PID = 0 bit stream. Note again that PIDs may have to be reassigned in this case.

Note also that in all descriptions of the higher level

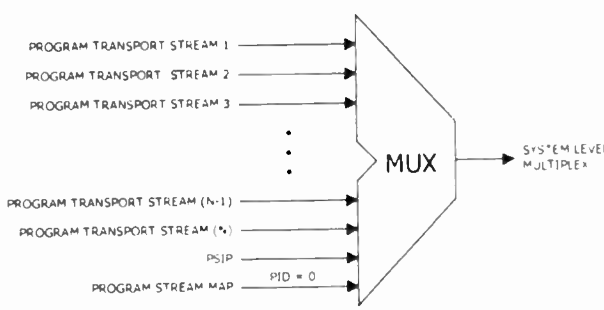


Figure 1.11-14. Illustration of the multiplex function to form the system level bit stream.

[10] The terminology can be confusing. The term **program** is analogous to a channel in NTSC; the term **program stream** refers to a particular bit stream format defined by MPEG but not used in the ATSC Digital Television Standard. **Program transport stream** is the term used to describe a transport bit stream that has been generated for a program.

multiplexing functionality, no mention is made of the functioning of the multiplexer and multiplexing policy that should be used. This function is not a part of the Standard and is up to individual designers. Because its basic function is one of filtering, the transport de-multiplexer will function on any digital television bit stream regardless of the multiplexing algorithm used.

RF/TRANSMISSION SYSTEMS

The RF/transmission subsystem portion of the DTV system is referred to as the VSB (for vestigial sideband) system. The VSB system offers two modes: a terrestrial broadcast mode, and a high data rate mode. The two modes share the same pilot, symbol rate, data frame structure, interleaving, Reed-Solomon (RS) coding and synchronization pulses. The terrestrial broadcast mode is optimized for maximum service area, and supports one ATSC DTV signal in a 6 MHz channel, with a payload data rate of 19.28... Mbps.

Both modes of the VSB transmission subsystem take advantage of a pilot, a segment sync and a training sequence for robust acquisition and operation. The two system modes also share identical carrier, sync, and clock recovery circuits, as well as phase correctors and equalizers. Additionally, both modes use the same RS code for forward error correction (FEC).

In order to maximize service area, the terrestrial broadcast mode incorporates both an NTSC rejection filter (in the receiver) and trellis coding. Pre-coding at the transmitter is incorporated in the trellis code. When the NTSC rejection filter is activated in the receiver, the trellis decoder is switched to a trellis code corresponding to the encoder trellis code concatenated with the filter.

The high data rate mode, on the other hand, does not have as severe an environment to work in as that of the terrestrial system (a cable environment is assumed). Therefore, a higher data rate is transmitted in the form of more data levels (bits/symbol). No trellis coding or NTSC interference rejection filters are employed.

VSB transmission inherently requires only processing the in-phase (I) channel signal, sampled at the symbol rate, thus optimizing the receiver for low cost implementation. The decoder only requires one A/D converter and a real (not complex) equalizer operating at the symbol rate of 10.76 Msamples/s.

The parameters for the two VSB transmission modes are shown in Table 1.11-4.

A functional block diagram of a representative 8 VSB terrestrial broadcast transmitter is shown in Figure 1.11-15. The input to the transmission subsystem from the transport subsystem is a 19.39.. Mbps serial data stream comprised of 188-byte MPEG-compatible data packets (including a sync byte and 187 bytes of data that represent a payload data rate of 19.28... Mbps).

The incoming data is randomized and then processed for FEC in the form of RS coding (20 RS parity bytes are added to each packet), 1/6 data field interleaving and 2/3-rate trellis coding. The randonization and FEC

Table 1.11-4
Parameters for VSB Transmission Modes

Parameter	Terrestrial mode	High data rate mode
Channel bandwidth	6 MHz	6 MHz
Excess bandwidth	11.5%	11.5%
Symbol rate	10.76 Msymbols/s	10.76 Msymbols/s
Bits per symbol	3	4
Trellis FEC	2/3 rate	None
Reed-Solomon FEC	T = 10 (207,187)	T = 10 (207,187)
Segment length	832 symbols	832 symbols
Segment sync	4 symbols per segment	4 symbols per segment
Frame sync	1 per 313 segments	1 per 313 segments
Payload data rate	19.28 Mbps	38.57 Mbps
NTSC co-channel rejection	NTSC rejection filter in receiver	N/A
Pilot power contribution	0.3 dB	0.3 dB
C/N threshold	14.9 dB[11]	28.3 dB

(Source: ATSC Doc A/54)

processes are not applied to the sync byte of the transport packet, which is represented in transmission by a *Data Segment Sync* signal. Following randomization and forward error correction processing, the data packets are formatted into Data Frames for transmission and Data Segment Sync and *Data Field Sync* are added.

Figure 1.11-16 shows how the data are organized for transmission. Each *Data Frame* consists of two *Data Fields*, each containing 313 *Data Segments*. The first Data Segment of each Data Field is a unique synchronizing signal (Data Field Sync) and includes the training sequence used by the equalizer in the receiver. The remaining 312 Data Segments each carry the equivalent of the data from one 188 byte transport packet plus its associated FEC overhead. The actual data in each Data Segment comes from several transport packets because of data interleaving. Each Data Segment consists of 832 symbols. The first four symbols are transmitted in binary form and provide segment synchronization. This Data Segment Sync signal also represents the sync byte of the 188 byte MPEG compatible transport packet. The remaining 828 symbols of each Data Segment carry data equivalent to the remaining 187 bytes of a transport packet and its associated FEC overhead. These 828 symbols are transmitted as 8-level signals and therefore carry three bits per symbol. Thus, 828 x 3 = 2484 bits of data are carried in each Data Segment, which exactly matches the requirement to send a protected transport packet:

187 data bytes + 20 RS parity bytes = 207 bytes
207 bytes $\times$ 8 bits/byte = 1656 bits
2/3 rate trellis coding requires 3/2 $\times$ 1656 bits = 2484 bits.

[11] Note that this published value was not confirmed during testing. The value used for noise limited service by the FCC was 15.19 dB, which was the test result selected by ACATS under one of many different test conditions measured by the ATTC.

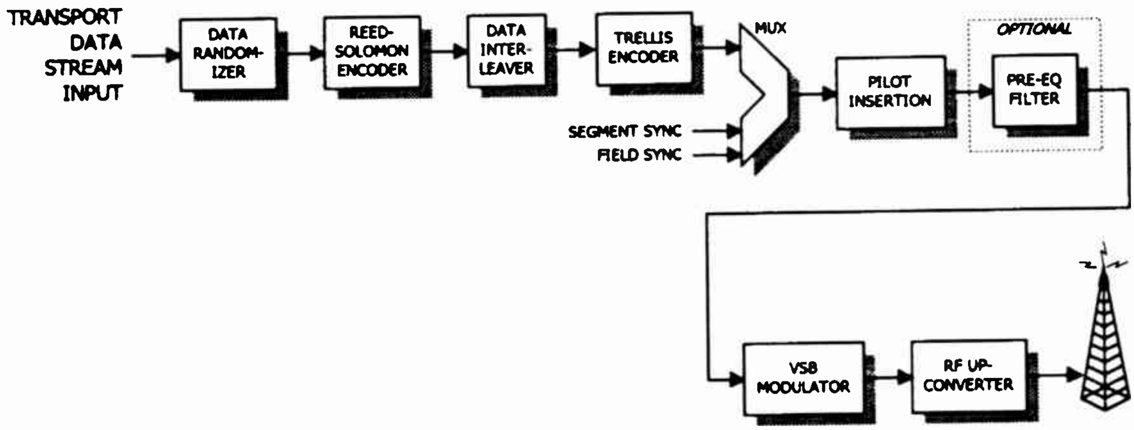


Figure 1.11-15. VSB transmitter.

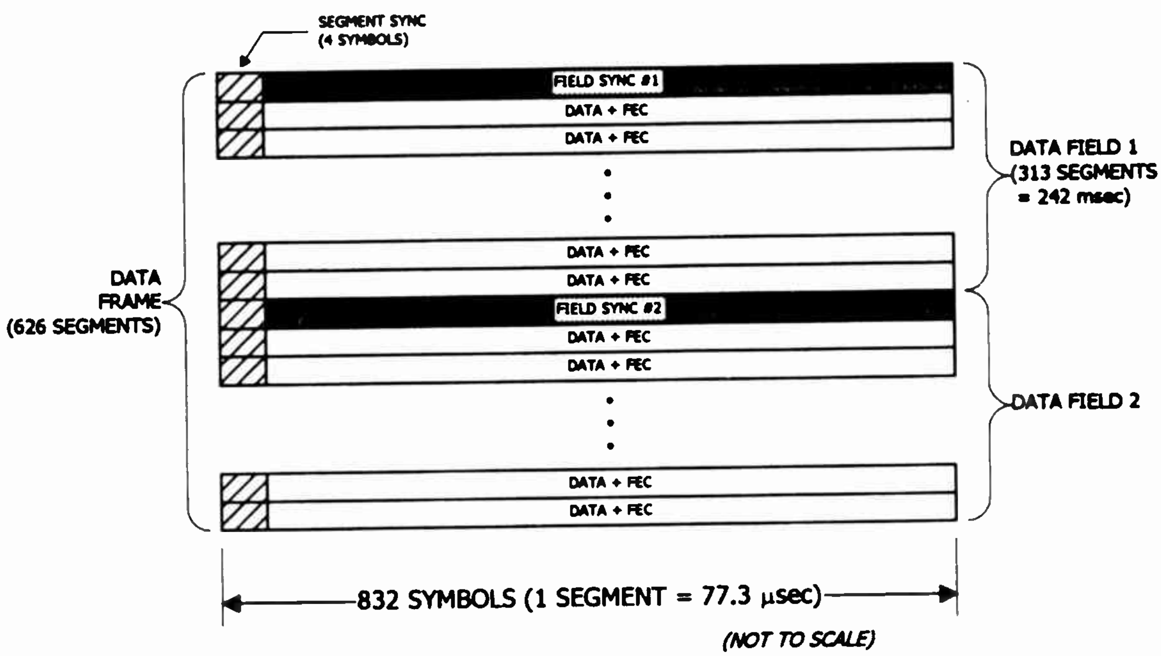


Figure 1.11-16. VSB data frame.

The exact symbol rate is given by:

$$S_r \text{ (MHz)} = 4.5/286 \times 684 = 10.76 \ldots \text{ MHz} \quad (1)$$

The frequency of a Data Segment is given by:

$$f_{seg} = S_r/832 = 12.94 \ldots \times 10^3 \text{ Data Segments/s} \quad (2)$$

The Data Frame rate is given by:

$$f_{frame} = f_{seg}/626 = 20.66 \ldots \text{ frames/s} \quad (3)$$

The symbol rate S_r and the transport rate T_r shall be locked to each other in frequency, where:

$$T_r = 2 \times \left(\frac{188}{208}\right)\left(\frac{312}{313}\right)\left(\frac{684}{286}\right) \times 4.5 = 19.39 \ldots \text{ Mbps}$$

and

$$S_r = \left(\frac{684}{286}\right) \times 4.5 = 10.76 \ldots \text{ Msps}$$

is the symbol rate S_r in Msymbols per second for the transmission subsystem.

The symbol rate must be locked in frequency to the transport rate. The transmission subsystem carries 2 information bits per trellis-coded symbol (fundamentally, 8 VSB modulation is 3 bits per transmitted symbol, however, in this context a trellis-coded symbol means *without* the trellis overhead, resulting in the 2 bits-per-symbol figure just stated), so the gross payload is:

$$10.76 \ldots \times 2 = 21.52 \ldots \text{Mbps} \qquad (4)$$

To find the net payload delivered to a decoder it is necessary to adjust (4) for the overhead of the Data Segment Sync, Data Field Sync, and RS FEC. Upon doing this the net payload bit rate of the 8 VSB terrestrial transmission subsystem becomes:

$$21.52 \ldots \text{Mbps} \times 312/313 \times 828/832 \times 187/207 = 19.28 \ldots \text{Mbps} \qquad (5)$$

The factor of 312/313 accounts for the Data Field Sync overhead of one Data Segment per field. The factor of 828/832 accounts for the Data Segment Sync overhead of four symbol intervals per Data Segment, and the factor of 187/207 accounts for the RS FEC overhead of 20 bytes per Data Segment.

The calculation of the net payload bit rate of the high data rate mode is identical except that 16 VSB carries 4 bits per symbol. Therefore, the net bit rate is twice that of the 8 VSB terrestrial mode:

$$19.28 \ldots \text{Mbps} \times 2 = 38.57 \ldots \text{Mbps} \qquad (6)$$

To get the net bit rate seen by a transport decoder, however, it is necessary to account for the fact that the MPEG sync bytes are removed from the data stream input to the 8 VSB transmitter. This amounts to the removal of one byte per data segment. These MPEG sync bytes are then reconstituted at the output of the 8 VSB receiver. The net bit rate seen by the transport decoder is:

$$19.28 \ldots \text{Mbps} \times 188/187 = 19.39 \ldots \text{Mbps} \qquad (7)$$

The net bit rate seen by the transport decoder for the high data rate mode is:

$$19.39 \ldots \text{Mbps} \times 2 = 38.78 \ldots \text{Mbps} \qquad (8)$$

The 8-level symbols combined with the binary Data Segment Sync and Data Field Sync signals are used to suppressed-carrier modulate a single carrier. Before transmission, however, most of the lower sideband is removed. The resulting spectrum is flat, except for the band edges where a nominal square root raised cosine response results in 620 kHz-wide transition regions.

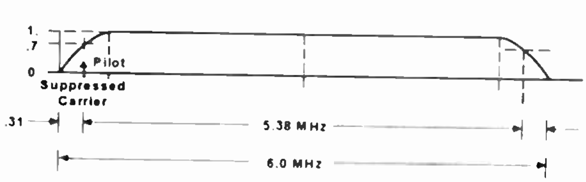


Figure 1.11-17. VSB channel occupancy (nominal) *(Source: ATSC Doc A/54).*

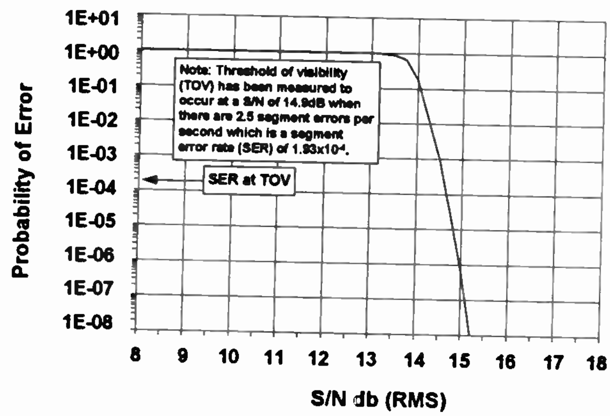


Figure 1.11-18. Segment error probability, 8 VSB with 4-state Trellis, RS (207,187) *(Source: ATSC Doc A/54).*

The nominal VSB transmission spectrum is shown in Figure 1.11-17. At the suppressed-carrier frequency, 309.441 kHz from the lower band edge, a small pilot is added to the signal.

The theoretical analysis of the terrestrial VSB system predicts that it can operate in a signal-to-additive-white-Gaussian-noise (S/N) environment of 14.9 dB. The 8 VSB, 4-state segment error probability curve in Figure 1.11-18 shows a segment error probability of $1.93 \times 10_{-4}$. This is equivalent to 2.5 segment errors/second which has been established by measurement as the threshold of visibility (TOV) of errors. Tests at ATTC showed a S/N of 15.19 dB with a BER of $\sim 1 \times 10_{-6}$ for the TOV in the strong signal, random noise test. Subsequently, the FCC elected to use 15.19 dB as the planning factor for the noise-limited service contour.

The cumulative distribution function (CDF) of the peak-to-average power ratio, as measured on a low power transmitted signal with no non-linearities, is plotted in Figure 1.11-19. This theoretical plot shows

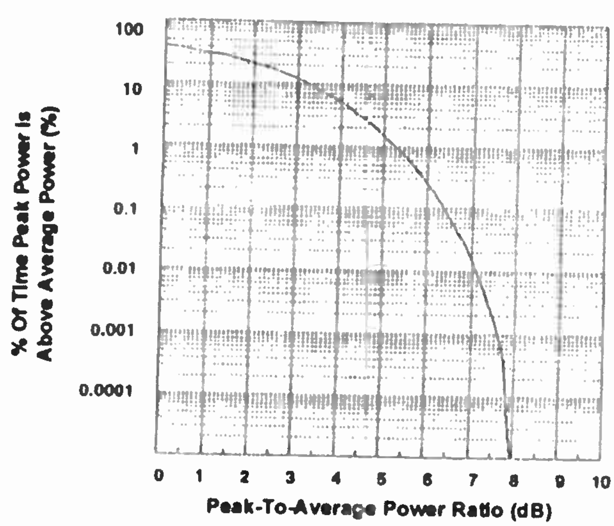


Figure 1.11-19. Theoretical cumulative distribution function of 8 VSB peak-to-average power ratio *(Source: ATSC Doc A/54).*

that 99.9% of the time the transient peak power should be within 6.3 dB of the average power. Tests of the Grand Alliance implementation showed the same shape curve, this result being further confirmed by the fact that two different measurement approaches yielded only slightly differing results. The 99.9% ratio was either 5.9 dB or 6.17 dB. The 99.99% of the time ratios were 6.6 and 6.89 dB. This is an important parameter because if the peak-to-average ratio of the transmitter is exceeded, clipping will occur as the signal transitions from one state to another. The effect of this clipping is generally expected to be negligible as long as the clipping is small, but this remains an area where long-term experience is needed to fully understand any impact. Although precise ratios appear in the test reports and standards documents, practical implementations of the DTV system cannot be expected to produce these exact performance levels.

A pre-equalizer filter is recommended for over-the-air broadcasts where the high power transmitter used may have significant in-band ripple or roll-off at band edges. This linear distortion can be detected by an equalizer in a reference demodulator (ideal receiver) located at the transmitter site that is receiving a small sample of the antenna signal feed provided by a directional coupler, recommended to be located at the sending end of the antenna feed transmission line. The resulting equalizer tap weights are then transferred into the transmitter pre-equalizer for pre-correction of transmitter"s linear distortion.

The VSB transmitter applies a two-step modulation process. In the first step, the baseband 8 VSB signal is modulated onto an IF carrier which will typically be the same for each transmission site regardless of the final RF frequency. Then, an RF upconverter translates the filtered, flat IF data signal spectrum to the desired RF channel. For the same approximate coverage as an NTSC transmitter (at the same frequency), the average power required of the ATSC DTV signal is approximately 12 dB less than the corresponding NTSC peak sync power for the same reception assumption.

The frequency of the RF upconverter oscillator in ATSC DTV terrestrial broadcasts will typically be the same as that used for an NTSC signal at the same channel (except for NTSC offsets). However, in extreme co-channel interference situations, the ATSC DTV system is designed to take advantage of precise RF carrier frequency offsets with respect to an NTSC co-channel carrier. As the VSB data signal sends repetitive synchronizing information (segment syncs), precise offset causes NTSC co-channel carrier interference into the VSB receiver to phase alternate from sync to sync. The VSB receiver circuits average successive syncs to cancel the interference and make data segment sync detection more reliable.

For ATSC DTV co-channel interference into NTSC, the interference is noise-like and does not change with precise offset. Even the ATSC DTV pilot interference into NTSC does not benefit from precise frequency offset because it is so small (11.3 dB below the data power) and falls far down the Nyquist slope (20 dB or more) of NTSC receivers.

Interference to an NTSC signal from an upper adjacent DTV channel is minimized by requiring (as part of the FCC rules) that the pilot carrier of DTV transmitters located within 88 kilometers of a lower-adjacent NTSC transmitter must be maintained at a precision frequency offset of 5.082138 MHz (±3 Hz with respect to that NTSC video carrier.

For DTV-into-DTV co-channel interference, precise carrier offset prevents possible misconvergence of the adaptive equalizer. If perchance the two ATSC DTV Data Field Sync signals should fall within the same data segment time, the adaptive equalizer could misinterpret the interference as a ghost. Although the FCC did not specify this parameter, testing for DTV to DTV co-channel interference thresholds was done with a frequency offset of 19,403 Hz ±5Hz separation and the FCC's planning factors used were the ones from this test.

DTV RECEIVER

The DTV system design partitioned responsibility for overall performance between the transmitter and the receiver. All testing of the system was based on the Grand Alliance implementation of a receiver. However the DTV receiver is not covered by any specific standard. The information in this section is provided to provide an overview of how receivers may work.

The following descriptions were based upon the Grand Alliance HDTV System Specification and are specific to the hardware implementation of the Grand Alliance. Figure 1.11-20 shows the receiver block diagram of the VSB terrestrial broadcast transmission system. Descriptions of each block follow.

The tuner receives the 6 MHz signal (UHF or VHF) from the antenna, and is a high-side injection double-conversion type with a first IF frequency of 920 MHz. This selection of first IF frequency is high enough so that the input band-pass filter selectivity prevents the local oscillator (978-1723 MHz) from leaking out the tuner front end and interfering with other UHF channels, yet it is low enough for second harmonics of UHF channels (470-806 MHz) to fall above the first IF band-pass.

The tuner input has a band-pass filter that limits the frequency range to 50-810 MHz, rejecting all other non-television signals that may fall within the tuner"s image frequency range (beyond 920 MHz). In addition, a broadband tracking filter rejects other television signals, especially those much larger in signal power than the desired signal power.

A 10 dB gain, wideband RF amplifier increases the signal level into the first mixer, and is the dominant determining factor of receiver noise figure (7-9 dB over entire VHF, UHF, and cable bands). The FCC planning factors assumed 7 dB Nf for UHF. The system Nf for LVHF and HVHF was assumed to be 10 dB. It is clear that, for LVHF, the FCC assumed the environmental noise component of this total to be 5 dB.

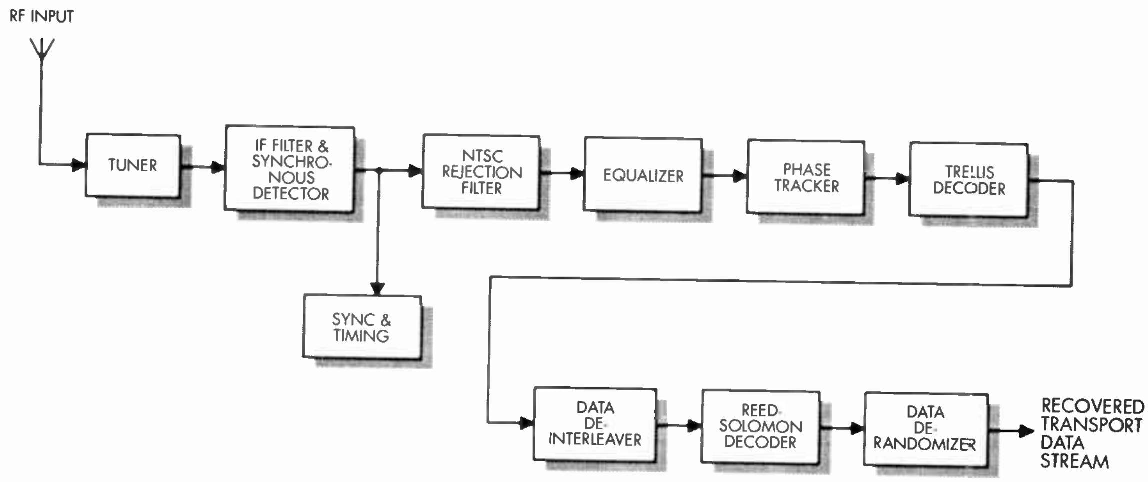


Figure 1.11-20. VSB receiver.

The assumptions about partitioning the system noise figure for HVHF were not disclosed. Interpolation of the receiver Nf from the two known values yields 5, 6, and 7 dB for LVHF, HVHF and UHF respectively. As of this writing no voluntary industry standards or requirements for receiver performance to match these values exist.

The first mixer is a linear double-balanced design to minimize even harmonic generation, is driven by a synthesized low phase noise local oscillator (LO) above the first IF frequency (high-side injection). This mixer is followed by an LC filter in tandem with a narrow 920 MHz band-pass ceramic resonator filter. A 920 MHz IF amplifier is placed between the two filters. Delayed AGC of the first IF signal is applied immediately following the first LC filter. The AGC circuit is intended to protect the remaining active stages from large signal overload.

The second mixer is driven by the second LO, which is an 876 MHz voltage-controlled SAW oscillator. It is controlled by the frequency and phase locked loop (FPLL) synchronous detector. The second mixer, whose output is at the desired 44 MHz second IF frequency, drives a constant gain 44 MHz amplifier. The output of the tuner feeds the IF SAW filter and synchronous detection circuitry.

Carrier recovery is performed on the small pilot carrier by an FPLL circuit. The first LO is synthesized by a PLL. The third LO is a fixed reference oscillator. Control for the second LO comes from the FPLL synchronous detector, which contains both a frequency loop and a phase locked loop in one circuit. The frequency loop provides a wide frequency pull-in range of ± 100 kHz while the phase locked loop has a narrow bandwidth (less than 2 kHz).

During frequency acquisition, the frequency loop uses both the in-phase (I) and quadrature-phase (Q) pilot signals. All other data processing circuits in the receiver use only the I channel signal. Prior to phase-lock, as is the condition after a channel change, the automatic frequency control (AFC) low-pass filter acts on the beat signal created by the frequency difference between the VCO and the incoming pilot.

When the frequency difference comes close to zero, the APC loop takes over and phase-locks the incoming IF signal to the third LO. The correct phase-lock polarity is determined by forcing the polarity of the pilot to positive. Once locked, only the phase-locked loop is active. The APC low-pass filter is wide enough to allow ± 100 kHz frequency pull-in, yet narrow enough to consistently reject all strong white noise (including other DTV signals) and NTSC co-channel interference signals. The PLL has a bandwidth that is narrow enough to reject most of the AM and PM generated by the data, yet is wide enough to track out any phase noise on the signal (and, hence, on the pilot) out to about 2 kHz.

The repetitive data segment syncs are detected from among the synchronously detected random data by a narrow bandwidth filter. From the data segment syncs, a properly phased 10.76 MHz symbol clock is created along with a coherent AGC control signal.

The 10.76 Msymbols/s ($684 \div 286 \times 4{,}500{,}000$ Hz) I channel composite baseband data signal (syncs and data) from the synchronous detector is converted by an A/D converter for digital processing. A PLL is used to derive a 10.76 MHz symbol clock for the receiver.

With the PLL free-running, the data segment sync detector containing a 4-symbol sync correlator looks for the two level syncs occurring at the specified repetition rate. The repetitive segment sync is detected while the random data is not, enabling the PLL to lock on the sampled sync from the A/D converter, and achieve data symbol clock synchronization. Upon reaching a predefined level of confidence (using a confidence counter) that the segment sync has been found, subsequent receiver loops are enabled.

Data Field Sync detection is achieved by comparing each received data segment from the A/D converter (after interference rejection filtering to minimize co-channel interference) with ideal field #1 and field #2

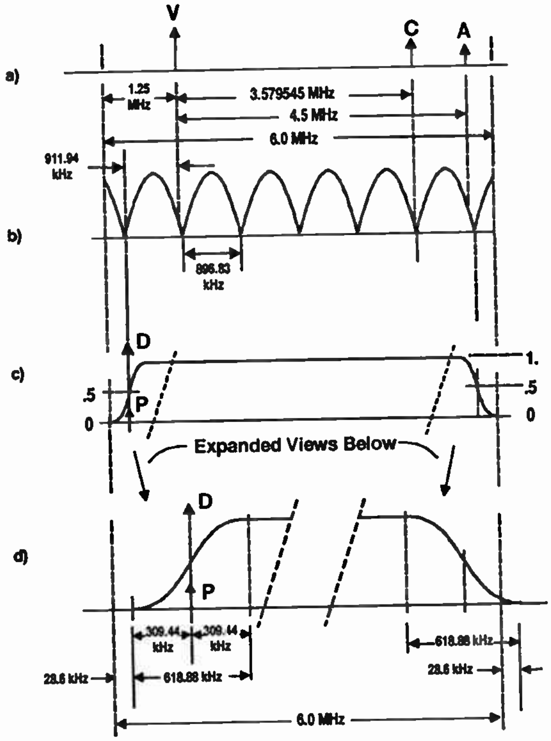


Figure 1.11-21. Location of NTSC carriers—comb filtering *(Source: ATSC Doc A/54).*

reference signals in the receiver. The field sync recovery circuit knows exactly where a valid field sync correlation should occur within each data segment, and only needs to perform a symbol-by-symbol difference.

NTSC Signal Rejection

The interference rejection properties of the VSB transmission system are based on the frequency location of the principal components of the NTSC co-channel interfering signal within the 6 MHz television channel and the periodic nulls of a VSB receiver baseband comb filter. Figure 1.11-21a shows the location and approximate magnitude of the three principal NTSC components: (1) the visual carrier (V) located 1.25 MHz from the lower band edge, (2) the chrominance subcarrier (C) located 3.58 MHz higher than the visual carrier frequency, and (3) the aural carrier (A) located 4.5 MHz higher than the visual carrier frequency.

The NTSC interference rejection filter (comb) is a one tap linear feed-forward filter, as shown in Figure 1.11-22. Figure 1.11-21b shows the frequency response of the comb filter, which provides periodic spectral nulls spaced 57 * f_H (10.762 MHz/12, or 896.85 kHz) apart. There are 7 nulls within the 6 MHz channel. The NTSC visual carrier frequency falls close to the second null from the lower band edge. The 6th null from the lower band edge is close to the NTSC chrominance subcarrier frequency, and the 7th null from the lower band edge is near the NTSC aural carrier.

Although the comb filter reduces the NTSC interference, the DTV datastream is also modified. The 7 data eyes (8 levels) are converted to 14 data eyes (15 levels). This conversion is caused by the partial response process which is a special case of intersymbol interference that does not close the data eye but creates double the number of eyes of the same magnitude. The trellis decoder can properly decode the modified DTV signal. Note that, because of sampling process, only the maximum data eye value is seen after A/D conversion.

NTSC interference can be detected by a circuit in which the signal-to-interference plus noise ratio of the binary Data Field Sync is measured at the input and the output of the comb filter, and then these values are compared. The comb filter is switched in and out of the system automatically, to achieve the lower interference energy.

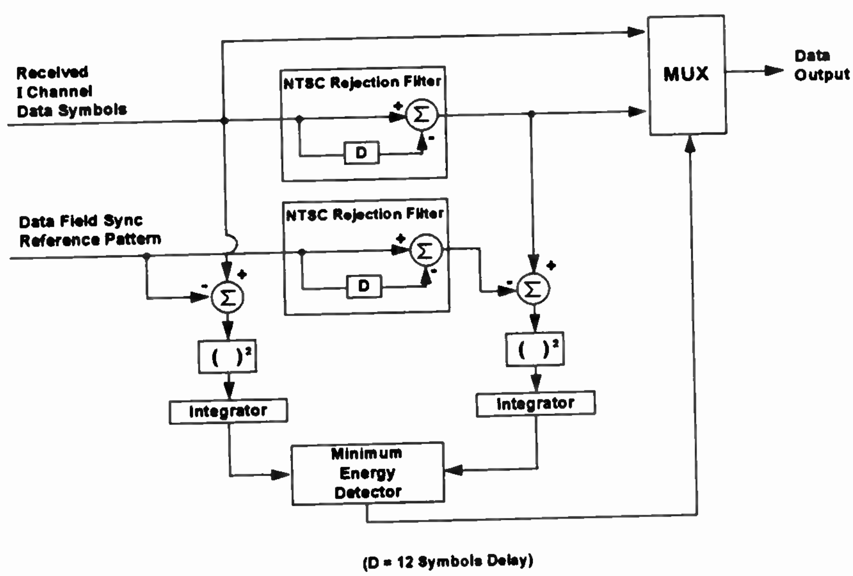


Figure 1.11-22. NTSC interference rejection filter *(Source: ATSC Doc A/54).*

There is a good reason to not leave the rejection comb filter switched in all the time. The comb filter, while providing needed co-channel interference benefits, degrades white noise performance by 3 dB. This is due to the fact that the filter output is the subtraction of two full-gain paths, and as white noise is uncorrelated from symbol to symbol, the noise power doubles. In addition, there is a further 0.5 dB of degradation due to the 12 symbol differential coding. When the NTSC service has been phased out, the comb filter can be omitted from the design.

The 8 VSB channel spectrum was illustrated in Figure 1.11- 17. This spectrum is fixed in frequency relative to the pilot carrier, but it may be slightly offset from the assigned 6 MHz channel, depending upon interference criteria. The FCC requires an offset with respect to a lower adjacent 6 MHz NTSC signal, such that the DTV pilot is 5.082138 MHz ± 3 Hz from the NTSC video carrier. This shifts the whole DTV spectrum up by 22.697 kHz. If the upper-adjacent signal is another DTV channel, its spectrum is expected to be shifted upward so that no spectral overlapping occurs. (Note that the FCC does not require this second shift). If the upper-adjacent signal is an NTSC signal, the shift of the DTV signal pushes part of the DTV spectrum out of its channel, but it is below the (RF equivalent of the) Nyquist slope of an NTSC receiver where there is high attenuation.

There are also offsets that have been determined to minimize interference to the DTV signal from a co-channel NTSC or DTV signal. The FCC does not require these offsets. The DTV pilot should be 911.944 kHz ± 1 kHz below a co-channel NTSC video carrier. This shifts the DTV spectrum up by 28.615 kHz. Finally, the DTV pilot should be 19.403 kHz ± 10 Hz from a co-channel DTV pilot carrier. Since there are three different interference minimizing offsets, there may be situations where they conflict. No guidance is available for that case.

When there is an NTSC neighbor, and the NTSC signal is currently assigned either a plus or minus 10 kHz offset, the absolute frequency of the DTV pilot would shift to maintain the offsets indicated above.

Equalizer

The equalizer/ghost canceller compensates for linear channel distortions, such as tilt and ghosts. These distortions can come from the transmission channel or from imperfect components within the receiver.

The equalizer uses a Least-Mean-Square (LMS) algorithm and can adapt on the transmitted binary Training Sequence as well as on the random data. The LMS algorithm computes how to adjust the filter taps in order to reduce the error present at the output of the equalizer. This algorithm can achieve equalization through three means: it can adapt on the binary training sequence, the data symbols throughout the frame when the eyes are open; or on data when the eyes are closed (blind equalization). The principal difference among these three methods is how the error estimate is generated.

For adapting on the training sequence, the training signal presents a fixed data pattern in the data stream. Because the data pattern is known, the exact error is generated by subtracting the training sequence from the output. The training sequence alone, however, may not be enough to track dynamic ghosts as these require tap adjustments more often than the training sequence is transmitted. Therefore, once equalization is achieved, the equalizer can switch to adapting on data symbols throughout the frame, and produce an accurate error estimate by slicing the data with an eight-level slicer and subtracting it from the output signal.

For fast dynamic ghosts (airplane flutter) it is necessary to use a blind equalization mode to aid in acquisition of the signal. Blind equalization models the multilevel signal as binary data signal plus noise, and the equalizer produces the error estimate by detecting the sign of the output signal and subtracting a (scaled) binary signal from the output to generate the error estimate.

To perform the LMS algorithm, the error estimate is multiplied by delayed copies of the signal. The delay depends upon which tap of the filter is being updated. This multiplication produces a cross-correction between the error signal and the data signal. The size of the correlation corresponds to the amplitude of the residual ghost present at the output of the equalizer and indicates how to adjust the tap to reduce the error at the output.

A block diagram of the equalizer is shown in Figure 1.11-23. The DC bias of the input signal is first removed by subtraction. The DC may be caused by circuit offsets, non-linearities or shifts in the pilot caused by ghosts. The DC offset is tracked by measuring the DC value of the training signal.

The equalizer filter consists of two parts, a 64 tap feed-forward transversal filter followed by a 192 tap decision feedback filter. The equalizer operates at the 10.762 MHz symbol rate (T-sampled equalizer).

The output of the forward filter and feedback filter are summed to produce the output. This output is sliced by either an eight level slicer (15 level slicer when the comb filter is used) or a binary slicer depending upon whether the data eyes are open or not. This sliced signal has the training signal and segment syncs reinserted as

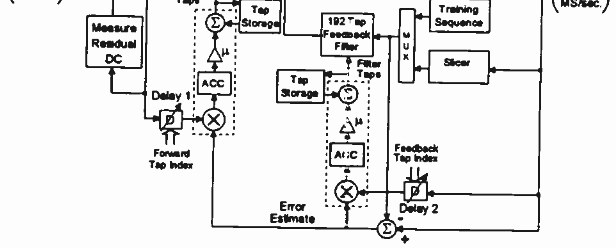


Figure 1.11-23. VSB receiver equalizer *(Source: ATSC Doc A/54)*.

these are fixed patterns of the signal. The resultant signal is fed into the feedback filter, and subtracted from the output signal to produce the error estimate. The error estimate is correlated with the input signal (for the forward filter), or by the output signal (for the feedback filter). This correlation is used to adjust the value of the tap. The delay setting of the adjustable delays is controlled according to the index of the filter tap that is being adjusted.

Phase Tracking Loop

The phase tracking loop is an additional decision feedback loop which further tracks out phase noise which has not been removed by the IF PLL operating on the pilot. Thus, phase noise is tracked out by not just one loop, but two concatenated loops. Because the system is already frequency locked to the pilot by the IF PLL (independent of the data), the phase tracking loop bandwidth is maximized for phase tracking by using a first order loop. As the phase tracker is operating on the 10.76 Msymbol/s data, the bandwidth of the phase tracking loop is fairly large, approximately 60 kHz.

Trellis Decoder

To help protect the trellis decoder against short burst interference, such as impulse noise or NTSC co-channel interference, 12 symbol code intrasegment interleaving is employed in the transmitter. As shown in Figure 1.11-24, the receiver uses 12 trellis decoders in parallel, where each trellis decoder sees every 12th symbol. This code interleaving has all the same burst noise benefits of a 12 symbol interleaver, but also minimizes the resulting code expansion (and hardware) when the NTSC rejection comb filter is active.

Before the 8 VSB signal can be processed by the trellis decoder it is necessary to suspend the Segment Sync. The Segment Sync is not trellis encoded at the transmitter.

The trellis decoder performs the task of slicing and convolutional decoding. It has two modes; one when the NTSC rejection filter is used to minimize NTSC co-channel, and the other when it is not used. The insertion of the NTSC rejection filter is determined automatically (before the equalizer), with this information passed to the trellis decoder. When there is little or no NTSC co-channel interference, the NTSC rejection filter is not used, and an optimal trellis decoder is used to decode the four state trellis encoded data. Serial bits are re-created in the same order in which they were created in the encoder.

When the NTSC rejection filter is employed, a trellis decoder optimized for this partial response channel is used. This optimal code requires eight states. In order to minimize the expansion of trellis states, two measures are taken: (1) special design of the trellis code, and (2) twelve-to-one interleaving of the trellis encoding. The interleaving, which corresponds exactly to the 12 symbol delay in the NTSC rejection filter, makes it so that each trellis decoder only sees a one-symbol delay NTSC rejection filter. By minimizing the delay stages seen by each trellis decoder, the expansion of states is also minimized.

Data De-interleaver

The convolutional de-interleaver performs the exact inverse function of the transmitter convolutional interleaver. Its 1/6 data field depth, and intersegment dispersion properties allow noise bursts lasting about 193 σs to be handled. The de-interleaver uses Data Field Sync for synchronizing to the first data byte of the data field. The convolutional de-interleaver is shown in Figure 1.11-25.

Reed-Solomon Decoder

The trellis decoded byte data is sent to the (207,187) t = 10 RS decoder, where it uses the 20 parity bytes to perform the byte-error correction on a segment-by-segment basis. Up to 10-byte errors per data segment are corrected by the RS decoder. Any burst errors created by impulse noise, NTSC co-channel interference, or trellis decoding errors, are greatly reduced by the combination of the interleaving and RS error correction.

Data Derandomizer

A Pseudo Random Sequence (PRS) randomizes the data at the transmitter. The derandomizer accepts the error-corrected data bytes from the RS decoder, and applies the same PRS randomizing code to the data.

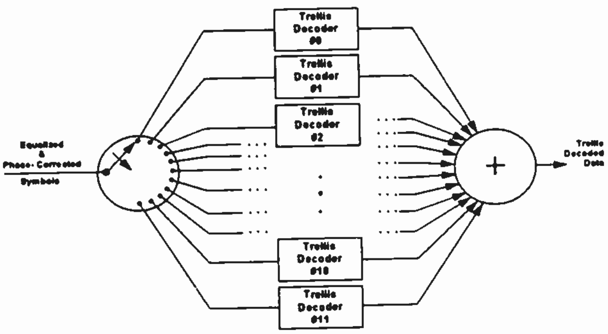


Figure 1.11-24. Trellis code de-interleaver *(Source: ATSC Doc A/54).*

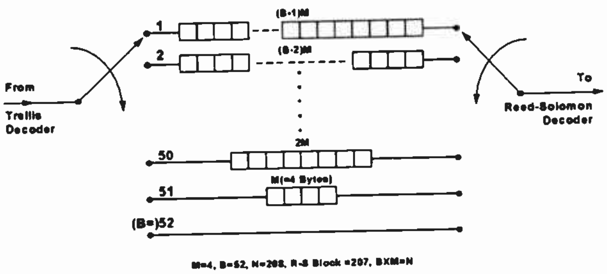


Figure 1.11-25. Convolutional de-interleaver *(Source: ATSC Doc A/54).*

The PRS code is generated identically as in the transmitter. Since the PRS is locked to the recovered Data Field Sync, it is exactly synchronized with the data.

PROGRAM AND SYSTEM INFORMATION PROTOCOL (PSIP)

The heart of the method to select one of the services in the DTV multiplex is a global assignment of a major_channel_number for each broadcaster and another number for each service which is locally controlled (minor_channel_number). The standard calls for the major channel to be the NTSC channel number of the station. The RF channel used for the (officially) minor facility change to add DTV broadcast capability is hidden from the consumers. The programs in the DTV multiplex are selected through a virtual channel mechanism, which the consumer sees as a second channel number (minor_channel_number). The DTV multiplex will carry program/service information about both the DTV programs (and data) and the current NTSC channel in a construct known as Program and System Information Protocol (PSIP) for terrestrial broadcast and cable.

The PSIP industry standard (ATSC document A/65) is a small collection of tables designed to operate within every transport stream for terrestrial broadcast of digital TV. It also contains tables for use by cable TV for the same purposes. It is needed so that consumers can select a TV program. PSIP is the result of combining and compacting the original two optional ATSC protocols: A/55 and A/56, whose implementation was difficult due to their structural differences and their overlapping definitions. PSIP solves this problem and establishes a channel tuning method that is uniform across terrestrial broadcast and cable. The tables defined in PSIP use packet identifiers (PIDs) that are different from those specified by the optional A/55 and A/56 standards that have in effect been superseded. The program paradigm is also superseded, and though its continued use is not precluded, its use does consume a significant portion of the PID number space.

An important feature of terrestrial broadcasting is that sources follow a distributed information model rather than a centralized one. Unlike cable or satellite, service providers are geographically distributed and have no interaction with respect to data unification or even synchronization. It was necessary to develop a protocol for describing system information and event descriptions that could be followed by every organization in charge of a physical transmission channel. System information allows navigation and access to each of the channels within the transport stream, whereas event descriptions give the user content information for browsing and selection. Meeting the needs of cable systems was a significant consideration, and both distribution systems requirements appear to be met. Work on extension of PSIP for satellite distribution has begun.

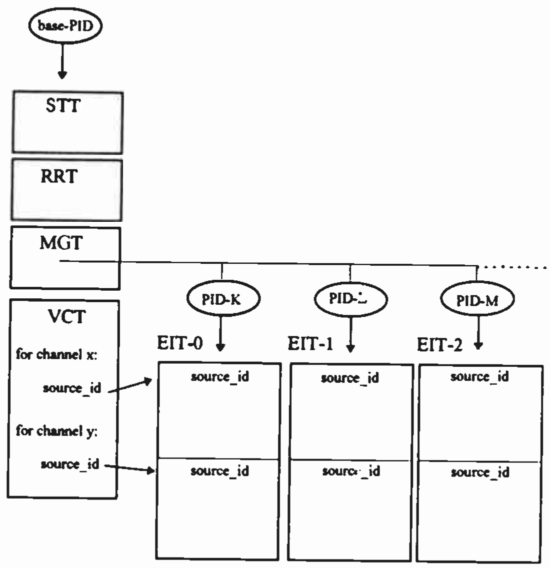


Figure 1.11-26. Main structure for the PSIP tables *(Source: ATSC Doc A/65).*

Elements of PSIP

PSIP is a collection of hierarchically-associated tables each of which describes particular elements of typical digital TV services. Figure 1.11-26 and Figure 1.11-27 show the different components and the notation used to describe them. The base tables are: the System Time Table (STT), the Rating Region Table (RRT), the Master Guide Table (MGT) and the Virtual Channel Table (VCT). A second set of tables includes the Event Information Tables (EIT) whose PIDs are defined in the MGT. A third set of tables includes the Extended Text Tables (ETT), and similarly, their PIDs are defined in the MGT.

The System Time Table (STT) is a small data structure that fits in one packet and serves as a reference for time of day. Receivers can use this table as a timing reference for start times of advertised events. System time for DTV is based on GPS and is a count of seconds since 12:00 a.m., January 6, 1980. The standard requires accuracy of ±4 seconds at the receiver, implying higher, but undefined, accuracy at the station.

The Rating Region Table (RRT) has been designed to communicate the definition of the selected

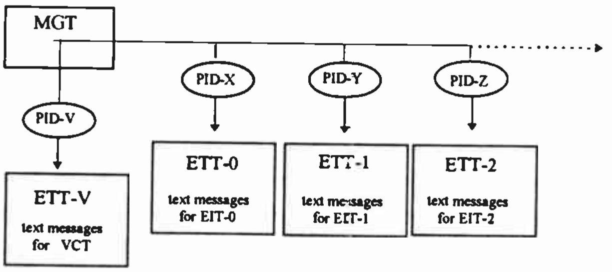


Figure 1.11-27. Extended text tables in the PSIP hierarchy. *(Source: ATSC Doc A/65)*

rating standard(s), such as the MPAA or the recently introduced television rating system. It has enough free space to include future developments, and it can also incorporate multiple multi-country aggregates definitions (to serve multiple countries or groups of countries simultaneously).

The Master Guide Table (MGT) provides general information about all of the other tables that comprise the PSIP standard. It defines table sizes necessary for memory allocation during decoding and version numbers to identify those tables that need to be updated. It also gives the PIDs which label the tables.

The Virtual Channel Table (VCT), also referred to as the Terrestrial VCT (TVCT), contains a list of all the channels that are or will be on-line, plus their attributes. Among the attributes are the channel name, navigation identifiers, stream components and types, etc.

As part of PSIP there are several EITs, each of which describes the events or TV programs associated with each of the virtual channels listed in the VCT. Each EIT is valid for a time interval of three hours. Since the total number of EITs is 128, up to 16 days of programming may be advertised in advance. EIT-0 always denotes the current three hours of programming, EIT-1 the next three hours, and so on. As a minimum, the first four EITs must always be present in every transport stream.

Start times for EITs are constrained to be one of the following Universal Coordinated Time (UTC) times: 0:00 (midnight), 3:00, 6:00, 9:00, 12:00 (noon), 15:00, 18:00, and 21:00. Imposing constraints on the start times as well as the interval duration is necessary for the purpose of re-multiplexing. During re-multiplexing, EIT tables coming from several distinct transport streams may end up grouped together or vice versa. If no constraints were imposed, re-multiplexing equipment would have to parse EITs by content in real time, which is a difficult task.

For example, consider a broadcast corporation operating in the Eastern time Zone of the U.S. This corporation decides to carry 6 EITs (18 hours of TV program information). If at present, the Eastern Time is 15:30 EDT (19:30 UTC), then the coverage times for the EIT tables are given in Table 1.11-5.

The abbreviation "nd" denotes next day. Before 17:00 EDT, the MGT will list the currently valid PIDs as: 123, 190, 237, 177, 295, and 221. At 17:00 EDT,

Table 1.11-5

An Example of EIT Coverage TimesTable

EIT number	Version Num.	Assigned PID	Coverage (UTC)	Coverage (EDT)
0	6	123	18:00–21:00	14:00–17:00
1	4	190	21:00–24:00	17:00–20:00
2	2	237	0:00–3:00	20:00–23:00
3	7	177	3:00–6:00	23:00–2:00 (nd)
4	8	295	6:00–9:00	2:00 (nd)–5:00 (nd)
5	15	221	9:00–12:00	5:00 (nd)–8:00 (nd)

(Source: ATSC Doc A/65)

table EIT-0 will become obsolete while the other ones will remain valid. At that time, the PID list can be changed to 190, 237, 177, 295, 221, maintaining the version number list as 4, 2, 7, 8, 15. Therefore, by simply shifting the listed PID values in the MGT, table EIT-1 can become EIT-0, table EIT-2 can become EIT-1, and so on.

However, it is also possible to regenerate one or several EITs at any time for correcting and/or updating the content (in cases where to be assigned events become known). Regeneration of EITs is flagged by updating version fields in the MGT. For example, if table EIT-2 needs to be updated at 16:17 EDT, then the new table must be transmitted with a version number equal to 3. Whenever the decoder monitoring the MGT detects a change in the version number of a table, it assumes that the table has changed and needs to be reloaded.

As illustrated in Figure 1.11-27, there can be several Extended Text Tables (ETTs), each of them having its PID defined in the MGT. Each EIT can have one ETT. Similarly, the VCT can have one ETT. As its name indicates, the purpose of an Extended Text Table (ETT) is to carry text messages. For example, for channels in the VCT, the messages can describe channel information, cost, coming attractions, etc. Similarly, for an event such as a movie listed in the EIT, the typical message is a short paragraph that describes the movie itself. ETTs are optional.

The minimum amount of information required in an ATSC terrestrial digital transport stream is the VCT, the MGT, the RRT, the STT, and the first four EITs. All of the other elements are optional.

PSIP Application Example

For the purpose of this example, assume that a broadcast station, here denominated NBZ, manages the frequency bands for RF channels 12 and 39. The first one is its analog channel whereas the second one will be used for digital broadcast. According to the premises established in A/65, NBZ must carry the PSIP tables in the digital transport stream of RF channel 39. The tables must describe TV programs and other services provided on RF channel 39 but can also describe information for the analog RF channel 12.

Assume that NBZ operates in the Eastern time zone of the U.S., and that the current time is 15:30 EDT (19:30 UTC). NBZ decides to operate in minimal configuration, therefore only the first four EITs need to be transmitted. As explained previously, EIT-0 must carry event information for the time window between 14:00 and 17:00 EDT, whereas EIT-1 to EIT-3 will cover the subsequent nine hours. For the first six hours, the scenarios presented in Table 1.11-6 and in Table 1.11-7 apply.

Similar tables can be built for the next six hours (for EIT-2 and EIT-3). According to this scenario, NBZ broadcasts four regular digital channels (also called virtual channels and denoted as VC), one matching the analog transmission (simulcast), another for sports, and a third one for movies. The fourth one

Table 1.11-6
The first 3-hour segment to be described in VCT and EIT-0

		14:00-14:30	14:30-15:00	5:00-15:30	115:30-16:00	16:00- 6:30	16:30-17:00
PTC 12	NBZ	City Life	City Life	Travel Show	Travel Show	News	News
PTC 39 VC #1	NBZ	City Life	City Life	Travel Show	Travel Show	News	News
PTC 39 VC #2	NBZ	Soccer	Golf Report	Golf Report	Car Racing	Car Racing	Car Racing
PTC 39 VC #3	NBZ	Secret Agent	Secret Agent	Lost Worlds	Lost Worlds	Lost Worlds	Lost Worlds
PTC 39 VC #4	NBZ	Headlines	Headlines	Headlines	Headlines	Headlines	Headlines

Table 1.11-7
The second 3-hour segment to be described in VCT and EIT-1

		17:00-17:30	17:30-18:00	18:00-18:30	18:30-19:00	19:00-19:30	19:30-20:00
PTC 12	NBZ	Music Today	NY Comedy	World View	World View	News	News
PTC 39 VC #1	NBZ	Music Today	NY Comedy	World View	World View	News	News
PTC 39 VC #2	NBZ	Car Racing	Car Racing	Sports News	Tennis Playoffs	Tennis Playoffs	Tennis Playoffs
PTC 39 VC #3	NBZ	Preview	The Bandit	The Bandit	The Bandit	The Bandit	Preview
PTC 39 VC #4	NBZ	Headlines	Headlines	Headlines	Headlines	Headlines	Headlines

Note: Physical Transmission Channel (PTC) (Source: ATSC Doc A/65)

supports a service displaying headlines with text and images.

The Master Guide Table (MGT)

The purpose of the MGT is to describe everything about the other tables, listing features such as version numbers, table sizes and PIDs. Figure 1.11-28 shows a typical MGT indicating, in this case, the existence in the transport stream of a VCT, the RRT, four EITs, one ETT for channels, and two ETTs for events.

The first entry of the MGT describes the version number and size of the VCT. The second entry corresponds to an instance of the RRT. If some region"s policy makers decided to use more than one instance of an RRT, the MGT would list each PID, version number, and size.

The next entries in the MGT correspond to the first four EITs that must be supplied in the transport stream. The user is free to choose their PIDs as long as they are unique in the MGT list of PIDs. After the EITs, the MGT indicates the existence of an ETT for channels carried using PID 0x1AA0. Similarly, the last two entries in the MGT signal the existence of two ETTs, one for EIT-0 and the other for EIT-1.

Descriptors can be added for each entry as well as for the entire MGT. By using descriptors, future improvements can be incorporated without modifying the basic structure of the MGT. The MGT is like a flag table that continuously informs the decoder about the status of all the other tables (except the STT which has an independent function). The MGT is continuously monitored at the receiver to prepare and anticipate changes in the channel/event structure. When tables are changed at the broadcast side, their version numbers are incremented and the new numbers are listed in the MGT. Based on the version updates and on the memory requirements, the decoder can reload the newly defined tables for proper operation.

MGT

table_type	PID	version num.	table size
VCT	0x1FFB (base_PID)	4	485 bytes
RRT - USA	0x1FFB (base_PID)	1	560 bytes
EIT-0	0x1FD0	6	2730 bytes
EIT-1	0x1FD1	4	1342 bytes
EIT-2	0x1DD1	2	1224 bytes
EIT-3	0x1DB3	7	1382 bytes
ETT for VCT	0x1AA0	21	4232 bytes
ETT-0	0x1BA0	10	32420 bytes
ETT-1	0x1BA1	2	42734 bytes

Figure 1.11-28. Content of the master guide table. *(Source: ATSC Doc A/65)*

The Virtual Channel Table (VCT)

Figure 1.11-29 shows the structure of the VCT that essentially contains the list of channels available in the transport stream. For convenience, it is possible to include analog channels and even other digital channels found in different transport streams.

The field number_channels_in_section indicates the number of channels described in one section of the VCT. In normal applications, as in the example being considered here, all channel information will fit into one section. However, there may be rare times when most of the physical channel is used to convey dozens of low-bandwidth services such as audio only and data channels in addition to one video program. In those cases, the channel information may be larger than the VCT section limit of 1 kB and therefore VCT segmentation will be required.

VCT

current_next_indicator = 1
number_channels_in_section = 5

major num.	minor num	short name	carrier freq. (MHz)	channel TS ID	progr. num	flags	service type	source id	descriptors
12	0	NBZ	205.25	0x0AA0	0x0AA0	--	analog	20	ch_name
12	1	NBZD	620.31	0x0AA1	0x00F1	--	digital	21	ch_name serv_locat.
12	5	NBZ-S	620.31	0x0AA1	0x00F2	--	digital	2	ch_name serv_locat.
12	12	NBZ-M	620.31	0x0AA1	0x00F3	--	digital	23	ch_name serv_locat.
12	31	NBZ-H	620.31	0x0AA1	0x00F8	--	digital	24	ch_name serv_locat.

Figure 1.11-29. Content of the virtual channel table. *(Source: ATSC Doc A/65)*

For example, assuming that a physical channel conveys 20 low-bandwidth services in addition to a TV program, and assuming that their VCT information exceeds 1 kB, then two or more sections may be defined. The first section may describe 12 virtual channels and the second 9 if such a partition leads to VCT sections with less than 1 kB.

A new VCT containing updated information can be transmitted at any time with the version_number increased by one. However, since a VCT describes only those channels from a particular transport stream, virtual channels added to the VCT at arbitrary times will not be detected by the receiver until it is tuned to that particular transport stream. For this reason, it is highly recommended that channel addition be made in advance to give the receivers the opportunity to scan the frequencies and detect the channel presence.

The fields major_channel_number and minor_channel_number are used for identification. The first one, the major channel number, is used to group all channels that are to be identified as belonging to a particular broadcast corporation (or particular identifying number such as 12 in this case). The minor channel number specifies a particular channel within the group.

The field short_name is a seven character name for the channel and may allow text-based access and navigation. The fields transport_stream_id and program number are included to link the VCT with the PAT and PMT. A sequence of flags follows these fields. The flags indicate: (1) if the channel is hidden (for NVOD applications), (2) if the channel has a long text message in the VCT-ETT and (3) if the channel is visible in general or has some conditional access constraints.

After the flags, a description of the type of service offered is included, followed by the source_id. The source_id is simply an internal index for representing the particular logical channel. EITs and ETTs use this number to provide a list of associated events or text messages respectively.

Two descriptors are associated with the logical channels in the example. The first one is extended_channel_name and, as its name indicates, it gives the full name of the channel. An example for channel NBZ-S could be: "NBZ Sports and Fitness." The other one, the service_location descriptor, is used to list the available bit streams and their PIDs necessary to decode packets at the receiver. Assuming that NBZ-M offers bilingual transmission, then the following attributes are tabulated within its service_location descriptor:

PID_audio_1	AC-3 audio	English
PID_audio_2	AC-3 audio	Spanish
PID_video	MPEG-2 video	No lang

Another method would be to provide a separate minor channel and bit entry to advertise availability of the second language as a separate program. A maximum of two VCTs may exist simultaneously in a transport stream: the current and the next VCT. The current VCT is recognized by having the flag current_next_indicator set to 1, while the next one has this flag set to 0. Although carrying the next VCT is optional, its

EIT- 0
source_id = 22 (NBZ-S instance)

num_events_in_section = 3

event ID	start time	length (seconds)	ETM location	title	descriptors
51	12:30	7200	01	Soccer Live	content_advisory
52	14:30	3600	00	Golf Report	closed_caption
53	15:30	9000	01	Car Racing	content_advisory

Figure 1.11-30. Content of EIT-0 for NBZ-S. *(Source: ATSC Doc A/65)*

use is recommended to give receivers advance notification of the new parameters that become operational during a VCT update.

Assume for example that a transport stream contains a VCT with a version number of 6 which has been operational for 20 hours. At 10:00 p.m., a football game using much more bandwidth will be broadcast, and for this reason, the number of available channels and PIDs will be redefined. A few minutes before the change (10:00 p.m. in this example) simultaneous transmission of the next VCT can start with a version number of 7. By continuously monitoring the MGT, a receiver can be informed that a next VCT is available. The receiver may want to cache the new VCT for future use. The receiver continues monitoring the MGT and when this table signals a version change for the current VCT (from 6 to 7), then the cached information can be used.

When the VCT refers to an analog service type, the channel_TSID TSID cannot refer to the identifier of a transport stream in the MPEG-2 sense. Analog NTSC broadcast signals can, however, carry a 16 bit unique identifier called a *Transmission Signal Identifier.*[12] For the example VCT in Figure D.4, the Transmission Signal Identifier for channel 12.0 is 0x0AA0. A receiver can use the Transmission Signal ID given in the analog channel's channel_TSID field to verify that the NTSC signal received at the frequency given in the VCT is actually the desired signal. In the case that the Transmission Signal ID is not known or not available, the channel_TSID field may contain 0xFFFF to indicate "unknown."

The Event Information Tables (EITs)

The purpose of an EIT is to list all events for those channels that appear in the VCT for a given time window. As mentioned before, EIT-0 describes the events for the first 3 hours, EIT-1 for the next 3 hours, and so on. EIT-i and EIT-j have different PIDs as defined in the MGT. In PSIP, tables can have a multitude of instances. The different instances of a table share the same table_id value and PID but use different table_id_extension values.

In PSIP, an instance of EIT-k contains the list of events for a single virtual channel with a unique source_id. For this reason, the table_id_extension has been renamed as source_id in the EIT syntax. Figure 1.11-30 shows, for example, the NBZ-S instance for EIT-0. Following similar procedures, the NBZD, NBZ-M, and NBZ-H instances of EIT-0 can be constructed. The process can be extended and repeated to obtain all of the instances for the other tables in the time sequence: EIT-1, EIT-2, etc.

The three events programmed for the 3-hour period for NBZ-S are listed in Figure 1.11-30. The field event_id is a number used to identify each event. If an event time period extends over more than one EIT, the same event_id has to be used. The event_id is used to link events with their messages defined in the ETT, and therefore it has to be unique only within a virtual channel and a 3-hour interval defined by EITs. The

[12] EIA has standardized the Transmission Signal Identifier in EIA-752 *Transport of Transmission Signal Identifier (TSID) Using Extended Data Service (XDS).* This standard defines how to carry the analog TSID in a new XDS packet, within the Channel Information class.

event_id is followed by the start_time and then the length_in_seconds. Notice that events can have start times before the activation time of the table. The ETM_location specifies the existence and the location of an Extended Text Message (ETM) for this event. ETMs are simply long textual descriptions. The collection of ETMs constitutes an ETT.

An example of an ETM for the car racing event may be:

> Live coverage from Indianapolis. This car race has become the largest single-day sporting event in the world. Two hundred laps of full action and speed.

Several descriptors can be associated with each event. One of these is the content advisory descriptor that assigns a rating value according to one or more systems. Recall that the actual rating system definitions are tabulated within the RRT. When a closed caption descriptor is included, it signals the existence of closed captioning and lists the necessary parameters for decoding.

The details of the information carried in the closed captioning field are specified in EIA-708, *Advanced Television Closed Captioning*. EIA-708 also defines how to carry the current EIA-608 closed captions. It also defines a new set of captions which are based on a window concept with color, shading, backgrounds and positioning information to take advantage of the ten-fold increase in data rate allocated to closed captioning in DTV.

EIA-708 is intended primarily for receiver and captioning encoder manufacturers. It describes a layered protocol stack starting from the transport layer and moving up to an interpretation layer which generally corresponds to the application layer in the OSI stack. It contains a description of the transport method, the new data packets and their structures, specification of how the captioning data is to be processed, minimum implementation requirements for receivers, and a set of recommended practices for encoder and decoder manufacturers.

Note that EIA-708 was completed before A/65 was completed, so it referenced T3/S8-145, which was incorporated into A/65. The location of the caption service descriptor in the DTV multiplex changed after EIA-708 was approved, obsoleting some descriptive references, but not the functionality. From a broadcaster perspective, one should expect that program material may have either 608 style closed captions or 708 style captions, or both. It is expected that if 708 style captions are present then 608 style captions will typically also be present. Standards for how to transport/store this data in pre-encoded material have not been completed.

1.12

EAS STANDARD

DAVID E. WILSON
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC

INTRODUCTION

The primary purpose of the Emergency Alert System (EAS) is to provide a means for the President to communicate quickly with all citizens of the United States in the event of a national emergency. However, like its predecessor the Emergency Broadcast System (EBS), the EAS is most often activated for state, local and regional emergencies. Over-the-air broadcast stations provide the most important link in this system—the link that enables the emergency message to be simultaneously received by a large, yet locally targeted audience of both stationary and mobile individuals.

The FCC required EAS equipment to be installed and operational in all over-the-air broadcast stations beginning January 1, 1997. Cable systems that serve 10,000 or more subscribers must have EAS equipment installed and operational by December 31, 1998, and all others by October 1, 2002.

This chapter will analyze the EAS from four different perspectives. First, the modulation method used to generate the EAS signal will be discussed, including the four different tones and audio information that make up the EAS signal. Next, the coding structure used to create the digital message will be described. It is this digital coding that is the primary improvement of the EAS over the old EBS. Third, a description of how each party that participates in the EAS relates to each of the other parties participating in the system will be provided. Finally, the EAS responsibilities of individual broadcasters will be explained.

Modulation method

An EAS message consists of four elements in the following order: digital header code, two-tone attention signal, audio message describing the actual alert and digital end-of-message code. The components of an EAS message are illustrated in Figure 1.12-1. There are four different tones used to create an EAS message. Their frequencies are 853 Hz, 960 Hz, 1562.5 Hz and 2083.3 Hz, respectively. The first two tones are the same ones that were used to produce the well-known two-tone attention signal in the old EBS. They are used to produce the same two-tone signal in the EAS, although their duration is not as long as it was under the EBS—and they are used less frequently. The second two tones are used to produce the digital header and end-of-message codes, which are the principal improvements of the new EAS over the old EBS.

In the digital header and end-of-message codes the 2083.3 Hz tone represents a mark (or "1") and the 1562.5 Hz tone represents a space (or "0"). When a digital bit is transmitted, its associated tone is transmitted for precisely 1.92 msec, which is the amount of time it takes to transmit exactly three cycles of the 1562.5 Hz tone, and four cycles of the 2083.3 Hz tone. Figure 1.12-2 illustrates, in the time domain, how these tones can be merged together to create an 8-bit digital byte—or one ASCII character. The character depicted in Figure 1.12-2, for example, is the capital letter "U," as defined in ANSI X3.4-1977, which is the ASCII character set standard referenced in the FCC's EAS rules. In this form of ASCII text transmission, the least significant bit is transmitted first and the most significant (seventh) bit is transmitted last followed by an eighth null bit (either 0 or 1) to create an 8-bit byte. Thus, in Figure 1.12-2, the character being transmitted is "X1010101" which corresponds to the decimal ASCII character value of 85, and the capital letter "U."

The two-tone attention signal is produced by simultaneously transmitting an 853 Hz tone and a 960 Hz tone. Figure 1.12-3 is an illustration of the composite two-tone signal.

When the digital components of an EAS message are transmitted FCC rules require that they modulate the transmitter at no less than 80% of full channel modulation. When the two-tone attention signal is transmitted FCC rules require that each individual tone modulate the transmitter at no less than 40% of full

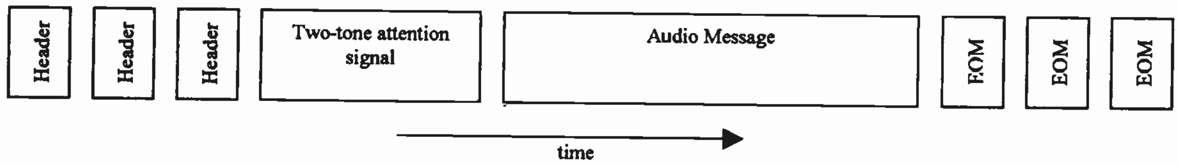


Figure 1.12-1. The primary elements of an EAS message.

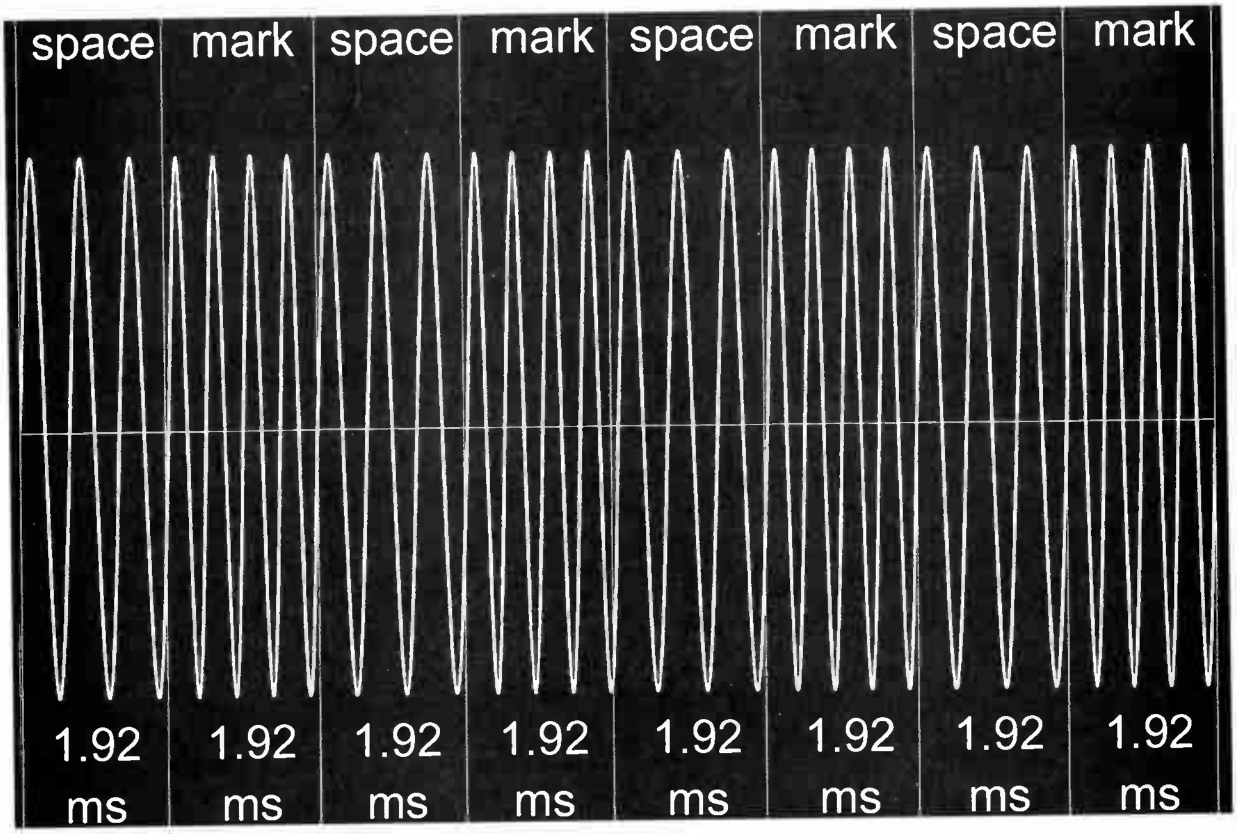


Figure 1.12-2. Time domain illustration of an EAS digital byte.

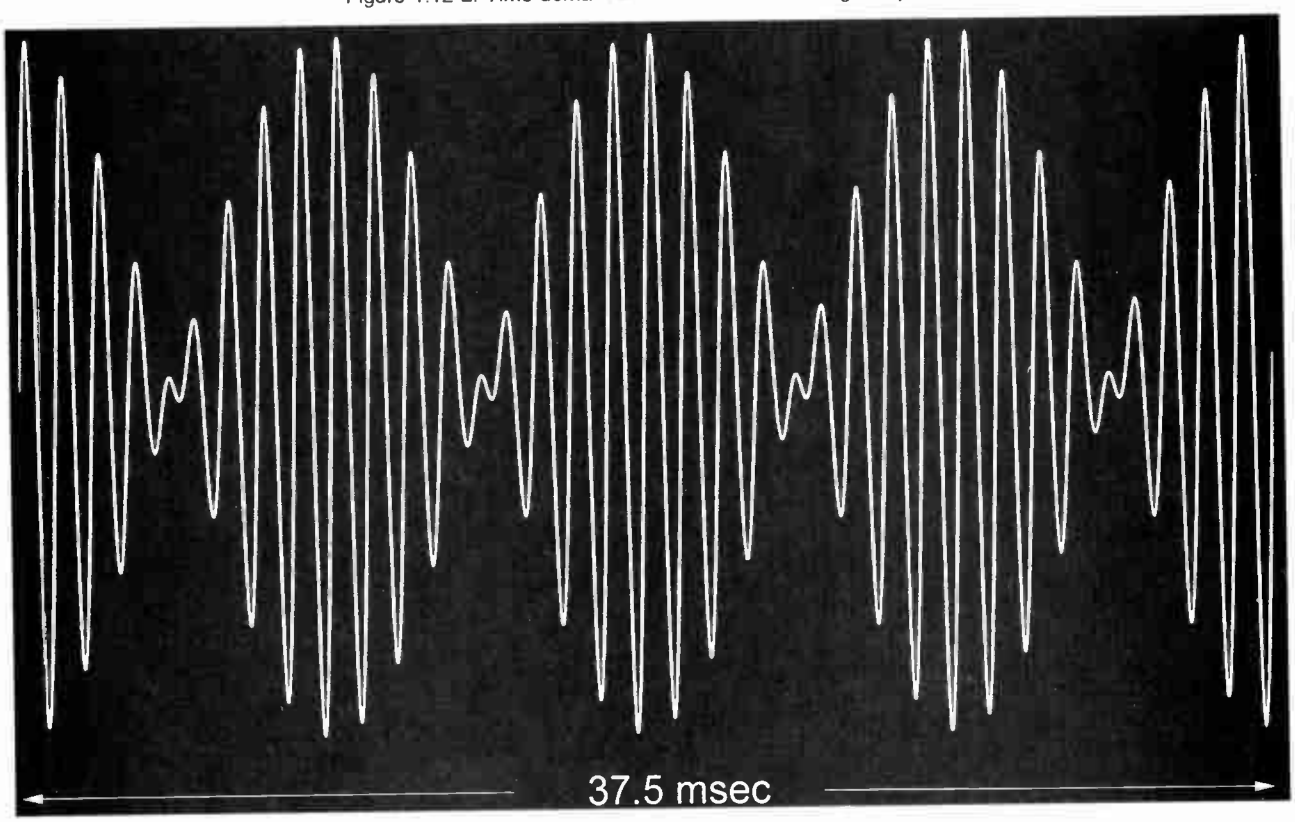


Figure 1.12-3. Time domain illustration of two-tone EAS attention signal.

channel modulation—and that the modulation levels of the two tones be within one dB of one another. When the audio portion of an EAS message is transmitted, it is subject to the same modulation requirements as the station's normal audio programming. This means that, for AM, FM and TV stations, the modulation level of the audio portion of an EAS message should be no less than 85% on peaks of frequent recurrence.

None of the components of an EAS message should exceed the maximum modulation limits specified for the relevant service. For AM stations, the maximum modulation level is 100% on negative peaks of frequent recurrence and 125% on positive peaks of frequent recurrence. For FM and TV stations the maximum modulation level is 100% on peaks of frequent recurrence—except that FM stations may increase their total modulation 0.5% for every 1% of subcarrier injections, up to a maximum level of 110% total modulation.

Coding Method

The header codes and the end-of-message codes in an EAS message are composed of a series of digital bytes and are transmitted using the 1562.5 Hz and 2083.3 Hz tones. The header and end-of-message codes are each repeated three times to ensure that they are received correctly. (For all EAS transmissions *except* the weekly test, television stations must transmit the header code data both aurally *and* visually. If they use a video crawl to do this, the crawl must be displayed at the top of the television screen or at another location where it will not interfere with other visual messages such as closed captioning.) The two-tone attention signal is created by simultaneously transmitting the 853 Hz and 960 Hz tones, and it is the same two-tone signal that was used in the old EBS. The actual audio message is simply an audio signal. The possibility exists that video and/or text messages may be transmitted in place of the audio message in the future, and television stations may, if they wish, transmit related video to accompany the audio message.

Except for the actual audio message itself the header is the portion of the EAS message that contains the most information. The two-tone attention signal is not intended to serve any other purpose than to audibly alert the audience that an EAS message is about to be broadcast, and the end-of-message code is simply a series a four ASCII "N" characters used to indicate that the alert is over and that the EAS equipment should reset itself to its normal, non-alert state. Both the header and end-of-message codes are each preceded by a preamble which is a string of 16 bytes (128 bits) consisting of ASCII character 171 ("$\frac{1}{2}$") repeated 16 times. The purpose of this preamble is to prepare EAS receivers to accept data by setting their automatic gain control (AGC) circuits and asynchronous clocking cycles.

Table 1.12-1
EAS Originator Codes

Originator	Originator Code
Broadcast station or cable system	EAS
Civil authorities	CIV
Emergency Action Notification Network	EAN
National Weather Service	WXR
Primary Entry Point System	PEP

The header code contains 42 bytes of information, which are illustrated in Figure 1.12-4. The first code segment in the header code consists of the characters "ZCZC" which indicates that an ASCII code sequence is about to begin. The next code segment is the Originator Code, which indicates who initiated activation of the EAS for the particular alert in question. The list of acceptable Originator Codes is provided in Table 1.12-1. The Event Code, which follows the Originator Code, indicates the type of emergency that is the subject of the alert. The Event Codes used in the EAS are the same ones defined in the Specific Area Message Encoder (SAME) standard used by the National Weather Service (see Table 1.12-2). Following the Event Code is the Location Code which indicates the state, county, and county subdivision (northwest, south central, etc.) that is affected by an EAS alert. The valid values for the three segments of the Location Code are described in Tables 1.12-3a, 3b and 3c. The FCC expects most alerts to have a Location Code that shows the alert to be effective over an entire county. It only expects the county subdivision feature to indicate something other than the entire county in cases involving oddly shaped or unusually large counties. The Valid Time Period of the alert, in HHMM format follows the Location Code. It indicates the length of time after the origination of the alert during which it is effective. The Valid Time Period of an EAS alert must be in 15 minute increments for periods that are no longer than an hour, and 30 minute increments for periods that are longer than an hour. The Origination Time of the alert follows the valid time period. It is in JJJHHMM format where JJJ represents the day of the year in Julian Calendar days (July 4 would be 185 in a non-leap year), and HHMM is in 24-hour format

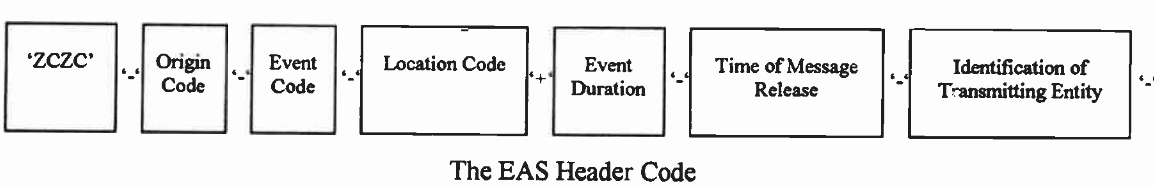


Figure 1.12-4. Components of the EAS Header Code.

Table 1.12-2
EAS Event Codes

Originator	Originator Code
National codes:	
Emergency Action Notification	EAN
Emergency Action Termination	EAT
National Information Center	NIC
National Periodic Test	NPT
Required Monthly Test	RMT
Required Weekly Test	RWT
Local codes:	
Administrative Message	ADR
Blizzard Warning	BZW
Civil Emergency Message	CEM
Evacuation Immediate	EVI
Flash Flood Statement	FFS
Flash Flood Warning	FFW
Flash Flood Watch	FFA
Flood Statement	FLS
Flood Warning	FLW
Flood Watch	FLA
High Wind Warning	HWW
High Wind Watch	HWA
Hurricane Statement	HLS
Hurricane Warning	HUW
Hurricane Watch	HUA
Practice/Demo Warning	DMO
Severe Thunderstorm Warning	SVR
Severe Thunderstorm Watch	SVA
Severe Weather Statement	SVS
Special Weather Statement	SPS
Tornado Warning	TOR
Tornado Watch	TOA
Tsunami Warning	TSW
Tsunami Watch	TSA
Winter Storm Warning	WSW
Winter Storm Watch	WSA

(10:00 PM is 22:00). Following the Origination Time is the Transmitting Station Identifier. This is the identification of the station transmitting the alert-not the station that originated the alert-but the station that is actually transmitting, or retransmitting it. When the EAS is activated, and an alert is making its way through the web network of stations, the Transmitting Station Identifier will be the only part of the header that gets changed as the message is relayed from one station to the next. All other components of the header will remain unchanged.

ASCII character 45, the dash/hyphen character (-), is used at the end of all but one of the code segments within the header code as a separator. The Location Code ends with ASCII character 43, the plus character (+).

Clearly, the EAS header code contains enough information to enable the receiving broadcast station to generate its own audible emergency message. Many stations choose to do this simply because they prefer to have their own announcers provide the emergency information to their audiences. During national level alerts and the required monthly test, however, stations are required to rebroadcast the audio that they receive without modification. One of the advantages of using the received audio for state and local emergencies is that, in some cases, the audio portion of the EAS message received by a station may contain more information than the header alone. For example, the header might indicate that there is a tornado watch for a particular county, but the audio portion of the received alert might indicate that a tornado has actually touched down at a particular location in the county.

SYSTEM ORGANIZATION

One of the primary improvements of the EAS over the old EBS is that it is generally more reliable since it utilizes a web structure instead of a daisy chain structure to connect all of the system's participants. In the EBS, each broadcast station was assigned one other station to monitor. In the EAS each station is

Table 1.12-3
EAS Location Codes

(a) State/Territory Codes

State/Territory	Code	State/Territory	Code	State/Territory	Code
Alabama	01	Maryland	24	South Carolina	45
Alaska	02	Massachusetts	25	South Dakota	46
Arizona	04	Michigan	26	Tennessee	47
Arkansas	05	Minnesota	27	Texas	48
California	06	Mississippi	28	Utah	49
Colorado	08	Missouri	29	Vermont	50
Connecticut	09	Montana	30	Virginia	51
Delaware	10	Nebraska	31	Washington	53
District of Columbia	11	Nevada	32	West Virginia	54
Florida	12	New Hampshire	33	Wisconsin	55
Georgia	13	New Jersey	34	Wyoming	56
Hawaii	15	New Mexico	35	American Samoa	60
Idaho	16	New York	36	Federated States of Micronesia	64
Illinois	17	North Carolina	37	Guam	66
Indiana	18	North Dakota	38	Marshall Islands	68
Iowa	19	Ohio	39	Northern Mariana Islands	69
Kansas	20	Oklahoma	40	Palau	70
Kentucky	21	Oregon	41	Puerto Rico	72
Louisiana	22	Pennsylvania	42	U.S. Minor Outlying Islands	74
Maine	23	Rhode Island	44	U.S. Virgin Islands	78

Table 1.12-3
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
AK	Aleutians East Borough	013
AK	Aleutians West Census Area	016
AK	Anchorage Borough	020
AK	Bethel Census Area	050
AK	Bristol Bay Borough	060
AK	Dillingham Census Area	070
AK	Fairbanks North Star Borough	090
AK	Haines Borough	100
AK	Juneau Borough	110
AK	Kenai Peninsula Borough	122
AK	Ketchikan Gateway Borough	130
AK	Kodiak Island Borough	150
AK	Lake and Peninsula Borough	164
AK	Matanuska-Susitna Borough	170
AK	Nome Census Area	180
AK	North Slope Borough	185
AK	Northwest Arctic Borough	188
AK	Prince of Wales-Outer Ketchikan Census	201
AK	Sitka Borough	220
AK	Skagway-Yakutat-Angoon Census Area	231
AK	Southeast Fairbanks Census Area	240
AK	Valdez-Cordova Census Area	261
AK	Wade Hampton Census Area	270
AK	Wrangell-Petersburg Census Area	280
AK	Yukon-Koyukuk Census Area	290
AL	Autauga County	001
AL	Baldwin County	003
AL	Barbour County	005
AL	Bibb County	007
AL	Blount County	009
AL	Bullock County	011
AL	Butler County	013
AL	Calhoun County	015
AL	Chambers County	017
AL	Cherokee County	019
AL	Chilton County	021
AL	Choctaw County	023
AL	Clarke County	025
AL	Clay County	027
AL	Cleburne County	029
AL	Coffee County	031
AL	Colbert County	033
AL	Conecuh County	035
AL	Coosa County	037
AL	Covington County	039
AL	Crenshaw County	041
AL	Cullman County	043
AL	Dale County	045
AL	Dallas County	047
AL	DeKalb County	049
AL	Elmore County	051
AL	Escambia County	053
AL	Etowah County	055
AL	Fayette County	057
AL	Franklin County	059
AL	Geneva County	061
AL	Greene County	063
AL	Hale County	065
AL	Henry County	067
AL	Houston County	069
AL	Jackson County	071
AL	Jefferson County	073
AL	Lamar County	075
AL	Lauderdale County	077
AL	Lawrence County	079
AL	Lee County	081
AL	Limestone County	083
AL	Lowndes County	085
AL	Macon County	087
AL	Madison County	089
AL	Marengo County	091
AL	Marion County	093
AL	Marshall County	095
AL	Mobile County	097
AL	Monroe County	099
AL	Montgomery County	101
AL	Morgan County	103
AL	Perry County	105
AL	Pickens County	107
AL	Pike County	109
AL	Randolph County	111
AL	Russell County	113
AL	St. Clair County	115
AL	Shelby County	117
AL	Sumter County	119
AL	Talladega County	121
AL	Tallapoosa County	123
AL	Tuscaloosa County	125
AL	Walker County	127
AL	Washington County	129
AL	Wilcox County	131
AL	Winston County	133
AR	Arkansas County	001
AR	Ashley County	003
AR	Baxter County	005
AR	Benton County	007
AR	Boone County	009
AR	Bradley County	011
AR	Calhoun County	013
AR	Carroll County	015
AR	Chicot County	017
AR	Clark County	019
AR	Clay County	021
AR	Cleburne County	023
AR	Cleveland County	025
AR	Columbia County	027
AR	Conway County	029
AR	Craighead County	031
AR	Crawford County	033
AR	Crittenden County	035
AR	Cross County	037
AR	Dallas County	039
AR	Desha County	041
AR	Drew County	043
AR	Faulkner County	045
AR	Franklin County	047
AR	Fulton County	049
AR	Garland County	051
AR	Grant County	053
AR	Greene County	055
AR	Hempstead County	057
AR	Hot Spring County	059
AR	Howard County	061
AR	Independence County	063
AR	Izard County	065
AR	Jackson County	067
AR	Jefferson County	069
AR	Johnson County	071
AR	Lafayette County	073
AR	Lawrence County	075
AR	Lee County	077
AR	Lincoln County	079
AR	Little River County	081
AR	Logan County	083
AR	Lonoke County	085
AR	Madison County	087
AR	Marion County	089
AR	Miller County	091
AR	Mississippi County	093
AR	Monroe County	095
AR	Montgomery County	097
AR	Nevada County	099
AR	Newton County	101
AR	Ouachita County	103
AR	Perry County	105
AR	Phillips County	107
AR	Pike County	109
AR	Poinsett County	111
AR	Polk County	113
AR	Pope County	115
AR	Prairie County	117
AR	Pulaski County	119
AR	Randolph County	121
AR	St. Francis County	123
AR	Saline County	125
AR	Scott County	127
AR	Searcy County	129
AR	Sebastian County	131
AR	Sevier County	133
AR	Sharp County	135
AR	Stone County	137
AR	Union County	139
AR	Van Buren County	141
AR	Washington County	143
AR	White County	145
AR	Woodruff County	147
AR	Yell County	149
AZ	Apache County	001
AZ	Cochise County	003
AZ	Coconino County	005
AZ	Gila County	007
AZ	Graham County	009
AZ	Greeniee County	011
AZ	La Paz County	012
AZ	Maricopa County	013
AZ	Mohave County	015
AZ	Navajo County	017
AZ	Pima County	019
AZ	Pinal County	021
AZ	Santa Cruz County	023
AZ	Yavapai County	025
AZ	Yuma County	027
CA	Alamada County	001
CA	Alpine County	003
CA	Amador County	005
CA	Butte County	007
CA	Calaveras County	009
CA	Colusa County	011
CA	Contra Costa County	013
CA	Del Norte County	015
CA	El Dorado County	017
CA	Fresno County	019
CA	Glenn County	021
CA	Humboldt County	023
CA	Imperial County	025
CA	Inyo County	027
CA	Kern County	029
CA	Kings County	031
CA	Lake County	033

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
CA	Lassen County	035
CA	Los Angeles County	037
CA	Madera County	039
CA	Marin County	041
CA	Mariposa County	043
CA	Mendocino County	045
CA	Merced County	047
CA	Modoc County	049
CA	Mono County	051
CA	Monterey County	053
CA	Nape County	055
CA	Nevada County	057
CA	Orange County	059
CA	Placer County	061
CA	Plumas County	063
CA	Riverside County	065
CA	Sacramento County	067
CA	San Benito County	069
CA	San Bernardino County	071
CA	San Diego County	073
CA	San Francisco County	075
CA	San Joaquin County	077
CA	San Luis Obispo County	079
CA	San Mateo County	081
CA	Santa Barbara County	083
CA	Santa Clara County	085
CA	Santa Cruz County	087
CA	Shasta County	089
CA	Sierra County	091
CA	Siskiyou County	093
CA	Solano County	095
CA	Sonoma County	097
CA	Stanislaus County	099
CA	Sutter County	101
CA	Tehama County	103
CA	Trinity County	105
CA	Tulare County	107
CA	Tuolumne County	109
CA	Ventura County	111
CA	Yolo County	113
CA	Yuba County	115
CO	Adams County	001
CO	Alamosa County	003
CO	Arapahoe County	005
CO	Archuleta County	007
CO	Baca County	009
CO	Bent County	011
CO	Boulder County	013
CO	Chaffee County	015
CO	Cheyenne County	017
CO	Clear Creek County	019
CO	Conejos County	021
CO	Costilla County	023
CO	Crowley County	025
CO	Custer County	027
CO	Delta County	029
CO	Denver County	031
CO	Dolores County	033
CO	Douglas County	035
CO	Eagle County	037
CO	Elbert County	039
CO	El Paso County	041
CO	Fremont County	043
CO	Garfield County	045
CO	Gilpin County	047
CO	Grand County	049
CO	Gunnison County	051
CO	Hinsdale County	053
CO	Huerfano County	055
CO	Jackson County	057
CO	Jefferson County	059
CO	Kiowa County	061
CO	Kit Carson County	063
CO	Lake County	065
CO	La Plata County	067
CO	Larimer County	069
CO	Las Animas County	071
CO	Lincoln County	073
CO	Logan County	075
CO	Mesa County	077
CO	Mineral County	079
CO	Moffat County	081
CO	Montezuma County	083
CO	Montrose County	085
CO	Morgan County	087
CO	Otero County	089
CO	Ouray County	091
CO	Park County	093
CO	Phillips County	095
CO	Pitkin County	097
CO	Prowers County	099
CO	Pueblo County	101
CO	Rio Blanco County	103
CO	Rio Grande County	105
CO	Routt County	107
CO	Saguache County	109
CO	San Juan County	111
CO	San Miguel County	113
CO	Sedgwick County	115
CO	Summit County	117
CO	Teller County	119
CO	Washington County	121
CO	Weld County	123
CO	Yuma County	125
CT	Fairfield County	001
CT	Hartford County	003
CT	Litchfield County	005
CT	Middlesex County	007
CT	New Haven County	009
CT	New London County	011
CT	Tolland County	013
CT	Windham County	015
DC	District of Columbia	001
DE	Kent County	001
DE	New Castle County	003
DE	Sussex County	005
FL	Alachua County	001
FL	Baker County	003
FL	Bay County	005
FL	Bradford County	007
FL	Brevard County	009
FL	Broward County	011
FL	Calhoun County	013
FL	Charlotte County	015
FL	Citrus County	017
FL	Clay County	019
FL	Collier County	021
FL	Columbia County	023
FL	Dade County	025
FL	DeSoto County	027
FL	Dixie County	029
FL	Duval County	031
FL	Escambia County	033
FL	Flagler County	035
FL	Franklin County	037
FL	Gadsden County	039
FL	Gilchrist County	041
FL	Glades County	043
FL	Gulf County	045
FL	Hamilton County	047
FL	Hardee County	049
FL	Hendry County	051
FL	Hernando County	053
FL	Highlands County	055
FL	Hillsborough County	057
FL	Holmes County	059
FL	Indian River County	061
FL	Jackson County	063
FL	Jefferson County	065
FL	Lafayette County	067
FL	Lake County	069
FL	Lee County	071
FL	Leon County	073
FL	Levy County	075
FL	Libery County	077
FL	Madison County	079
FL	Manatee County	081
FL	Marion County	083
FL	Martin County	085
FL	Monroe County	087
FL	Nassau County	089
FL	Okaloosa County	091
FL	Okeechobee County	093
FL	Orange County	095
FL	Osceola County	097
FL	Palm Beach County	099
FL	Pasco County	101
FL	Pinellas County	103
FL	Polk County	105
FL	Putnam County	107
FL	St. Johns County	109
FL	St. Lucie County	111
FL	Santa Rosa County	113
FL	Sarasota County	115
FL	Seminole County	117
FL	Sumter County	119
FL	Suwannee County	121
FL	Taylor County	123
FL	Union County	125
FL	Volusia County	127
FL	Wakulla County	129
FL	Walton County	131
FL	Washington County	133
GA	Appling County	001
GA	Atkinson County	003
GA	Bacon County	005
GA	Baker County	007
GA	Baldwin County	009
GA	Banks County	011
GA	Barrow County	013
GA	Bartow County	015
GA	Ben Hill County	017
GA	Berrien County	019
GA	Bibb County	021
GA	Bleckley County	023
GA	Brantley County	025
GA	Brooks County	027
GA	Bryan County	029
GA	Bulloch County	031
GA	Burke County	033
GA	Butts County	035
GA	Calhoun County	037
GA	Camden County	039
GA	Candler County	043
GA	Carroll County	045
GA	Catoosa County	047
GA	Charlton County	049
GA	Chatham County	051
GA	Chattahoochee County	053
GA	Chattooga County	055
GA	Cherokee County	057
GA	Clarke County	059
GA	Clay County	061

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
GA	Clayton County	063
GA	Clinch County	065
GA	Cobb County	067
GA	Coffee County	069
GA	Colquitt County	071
GA	Columbia County	073
GA	Cook County	075
GA	Coweta County	077
GA	Crawford County	079
GA	Crisp County	081
GA	Dade County	083
GA	Dawson County	085
GA	Decatur County	087
GA	DeKalb County	089
GA	Dodge County	091
GA	Dooly County	093
GA	Dougherty County	095
GA	Douglas County	097
GA	Early County	099
GA	Echols County	101
GA	Effingham County	103
GA	Elbert County	105
GA	Emanuel County	107
GA	Evans County	109
GA	Fannin County	111
GA	Fayette County	113
GA	Floyd County	115
GA	Forsyth County	117
GA	Franklin County	119
GA	Fulton County	121
GA	Gilmer County	123
GA	Glascock County	125
GA	Glynn County	127
GA	Gordon County	129
GA	Grady County	131
GA	Greene County	133
GA	Gwinnett County	135
GA	Habersham County	137
GA	Hall County	139
GA	Hancock County	141
GA	Haralson County	143
GA	Harris County	145
GA	Hart County	147
GA	Heard County	149
GA	Henry County	151
GA	Houston County	153
GA	Irwin County	155
GA	Jackson County	157
GA	Jesper County	159
GA	Jeff Davis County	161
GA	Jefferson County	163
GA	Jenkins County	165
GA	Johnson County	167
GA	Jones County	169
GA	Lamar County	171
GA	Lanier County	173
GA	Laurens County	175
GA	Lee County	177
GA	Liberty County	179
GA	Lincoln County	181
GA	Long County	183
GA	Lowndes County	185
GA	Lumpkin County	187
GA	McDuffie County	189
GA	McIntosh County	191
GA	Macon County	193
GA	Madison County	195
GA	Marion County	197
GA	Meriwether County	199
GA	Miller County	201
GA	Mitchell County	205
GA	Monrole County	207
GA	Montgomery County	209
GA	Morgan County	211
GA	Murray County	213
GA	Muscogee County	215
GA	Newton County	217
GA	Oconee County	219
GA	Oglethorpe County	221
GA	Paulding County	223
GA	Peach County	225
GA	Pickens County	227
GA	Pierce County	229
GA	Pike County	231
GA	Polk County	233
GA	Pulaski County	235
GA	Putnam County	237
GA	Quitman County	239
GA	Rabun County	241
GA	Randolph County	243
GA	Richmond County	245
GA	Rockdale County	247
GA	Schley County	249
GA	Screven County	251
GA	Seminole County	253
GA	Spalding County	255
GA	Stephens County	257
GA	Stewart County	259
GA	Sumter County	261
GA	Talbot County	263
GA	Taliaferro County	265
GA	Tattnall County	267
GA	Taylor County	269
GA	Telfair County	271
GA	Terrell County	273
GA	Thomas County	275
GA	Tift County	277
GA	Toombs County	279
GA	Towns County	281
GA	Treutlen County	283
GA	Troup County	285
GA	Turner County	287
GA	Twiggs County	289
GA	Union County	291
GA	Upson County	293
GA	Walker County	295
GA	Walton County	297
GA	Ware County	299
GA	Warren County	301
GA	Washington County	303
GA	Wayne County	305
GA	Webster County	307
GA	Wheeler County	309
GA	White County	311
GA	Whitfield County	313
GA	Wilcox County	315
GA	Wilkes County	317
GA	Wilkinson County	319
GA	Worth County	321
HI	Hawaii County	001
HI	Honolulu County	003
HI	Kalawao County	005
HI	Kauai County	007
HI	Maui County	009
IA	Adair County	001
IA	Adams County	003
IA	Allamakee County	005
IA	Appanoose County	007
IA	Audubon County	009
IA	Benton County	011
IA	Black Hawk County	013
IA	Boone County	015
IA	Bremer County	017
IA	Buchanan County	019
IA	Buena Vista County	021
IA	Butler County	023
IA	Calhoun County	025
IA	Carroll County	027
IA	Cass County	029
IA	Cedar County	031
IA	Cerro Gordo County	033
IA	Cherokee County	035
IA	Chickasaw County	037
IA	Clarke County	039
IA	Clay County County	041
IA	Clayton County	043
IA	Clinton County	045
IA	Crawford County	047
IA	Dallas County	049
IA	Davis County	051
IA	Decatur County	053
IA	Delaware County	055
IA	Des Moines County	057
IA	Dickinson County	059
IA	Dubuque County	061
IA	Emmet County	063
IA	Fayette County	065
IA	Floyd County	067
IA	Franklin County	069
IA	Fremont County	071
IA	Greene County	073
IA	Grundy County	075
IA	Guthrie County	077
IA	Hamilton County	079
IA	Hancock County	081
IA	Hardin County	083
IA	Harrison County	085
IA	Henry County	087
IA	Howard County	089
IA	Humboldt County	091
IA	Ida County	093
IA	Iowa County	095
IA	Jackson County	097
IA	Jasper County	099
IA	Jefferson County	101
IA	Johnson County	103
IA	Jones County	105
IA	Keokuk County	107
IA	Kossuth County	109
IA	Lee County	111
IA	Linn County	113
IA	Louise County	115
IA	Lucas County	117
IA	Lyon County	119
IA	Madison County	121
IA	Mahaska County	123
IA	Marion County	125
IA	Marshall County	127
IA	Mills County	129
IA	Mitchell County	131
IA	Monona County	133
IA	Monroe County	135
IA	Montgomery County	137
IA	Muscatine County	139
IA	O'Brien County	141
IA	Osceola County	143
IA	Page County	145
IA	Palo Alto County	147
IA	Plymouth County	149
IA	Pocahontas County	151
IA	Polk County	153
IA	Pottewattamie County	155
IA	Poweshiek County	157

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
IA	Ringgold County	159
IA	Sac County	161
IA	Scott County	163
IA	Shelby County	165
IA	Sioux County	167
IA	Story County	169
IA	Tama County	171
IA	Taylor County	173
IA	Union County	175
IA	Van Buren County	177
IA	Wapello County	179
IA	Warren County	181
IA	Washington County	183
IA	Wayne County	185
IA	Webster County	187
IA	Winnebago County	189
IA	Winneshiek County	191
IA	Woodbury County	193
IA	Worth County	195
IA	Wright County	197
ID	Ada County	001
ID	Adams County	003
ID	Bannock County	005
ID	Bear Lake County	007
ID	Benewah County	009
ID	Bingham County	011
ID	Blaine County	013
ID	Boise County	015
ID	Bonner County	017
ID	Bonneville County	019
ID	Boundary County	021
ID	Butte County	023
ID	Camas County	025
ID	Canyon County	027
ID	Caribou County	029
ID	Cassia County	031
ID	Clark County	033
ID	Clearwater County	035
ID	Custer County	037
ID	Elmore County	039
ID	Franklin County	041
ID	Fremont County	043
ID	Gem County	045
ID	Gooding County	047
ID	Idaho County	049
ID	Jefferson County	051
ID	Jerome County	053
ID	Kootenai County	055
ID	Latah County	057
ID	Lemhi County	059
ID	Lewis County	061
ID	Lincoln County	063
ID	Madison County	065
ID	Minidoka County	067
ID	Nez Perce County	069
ID	Oneida County	071
ID	Owyhee County	073
ID	Payette County	075
ID	Power County	077
ID	Shoshone County	079
ID	Teton County	081
ID	Twin Falls County	083
ID	Valley County	085
ID	Washington County	087
IL	Adams County	001
IL	Alexander County	003
IL	Bond County	005
IL	Boone County	007
IL	Brown County	009
IL	Bureau County	011
IL	Calhoun County	013
IL	Carroll County	015
IL	Cass County	017
IL	Champaign County	019
IL	Christian County	021
IL	Clark County	023
IL	Clay County	025
IL	Clinton County	027
IL	Coles County	029
IL	Cook County	031
IL	Crawford County	033
IL	Cumberland County	035
IL	DeKalb County	037
IL	De Witt County	039
IL	Douglas County	041
IL	DuPage County	043
IL	Edgar County	045
IL	Edwards County	047
IL	Effingham County	049
IL	Fayette County	051
IL	Ford County	053
IL	Franklin County	055
IL	Fulton County	057
IL	Gallatin County	059
IL	Greene County	061
IL	Grundy County	063
IL	Hamilton County	065
IL	Hancock County	067
IL	Hardin County	069
IL	Henderson County	071
IL	Henry County	073
IL	Iroquois County	075
IL	Jackson County	077
IL	Jasper County	079
IL	Jefferson County	081
IL	Jersey County	083
IL	Jo Deviess County	085
IL	Johnson County	087
IL	Kane County	089
IL	Kankakee County	091
IL	Kendall County	093
IL	Knox County	095
IL	Lake County	097
IL	La Salle County	099
IL	Lawrence County	101
IL	Lee County	103
IL	Livingson County	105
IL	Logan County	107
IL	McDonough County	109
IL	McHenry County	111
IL	McLean County	113
IL	Macon County	115
IL	Macoupin County	117
IL	Madison County	119
IL	Marion County	121
IL	Marshall County	123
IL	Mason County	125
IL	Massac County	127
IL	Menard County	129
IL	Mercer County	131
IL	Monroe County	133
IL	Montgomery County	135
IL	Morgan County	137
IL	Moultrie County	139
IL	Ogle County	141
IL	Peoria County	143
IL	Perry County	145
IL	Piatt County	147
IL	Pike County	149
IL	Pope County	151
IL	Pulaski County	153
IL	Putnam County	155
IL	Randolph County	157
IL	Richland County	159
IL	Rock Island County	161
IL	St. Clair County	163
IL	Saline County	165
IL	Sangamon County	167
IL	Schuyler County	169
IL	Scott County	171
IL	Shelby County	173
IL	Stark County	175
IL	Stephenson County	177
IL	Tazewell County	179
IL	Union County	181
IL	Vermilion County	183
IL	Wabash County	185
IL	Warren County	187
IL	Washington County	189
IL	Wayne County	191
IL	White County	193
IL	Whiteside County	195
IL	Will County	197
IL	Williamson County	199
IL	Winnebago County	201
IL	Woodford County	203
IN	Adams County	001
IN	Allen County	003
IN	Bartholomew County	005
IN	Benton County	007
IN	Blackford County	009
IN	Boone County	011
IN	Brown County	013
IN	Carroll County	015
IN	Cass County	017
IN	Clark County	019
IN	Clay County	021
IN	Clinton County	023
IN	Crawford County	025
IN	Daviess County	027
IN	Dearborn County	029
IN	Decatur County	031
IN	De Kalb County	033
IN	Delaware County	035
IN	Dubois County	037
IN	Elkhart County	039
IN	Fayette County	041
IN	Floyd County	043
IN	Fountain County	045
IN	Franklin County	047
IN	Fulton County	049
IN	Gibson County	051
IN	Grant County	053
IN	Greene County	055
IN	Hamilton County	057
IN	Hancock County	059
IN	Harrison County	061
IN	Hendricks County	063
IN	Henry County	065
IN	Howard County	067
IN	Huntington County	069
IN	Jackson County	071
IN	Jasper County	073
IN	Jay County	075
IN	Jefferson County	077
IN	Jennings County	079
IN	Johnson County	081
IN	Knox County	083
IN	Kosciusko County	085
IN	Lagrange County	087
IN	Lake County	089
IN	La Porte County	091
IN	Lawrence County	093

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
IN	Madison County	095
IN	Marion County	097
IN	Marshall County	099
IN	Martin County	101
IN	Miami County	103
IN	Monroe County	105
IN	Montgomery County	107
IN	Morgan County	109
IN	Newton County	111
IN	Noble County	113
IN	Ohio County	115
IN	Orange County	117
IN	Owen County	119
IN	Parke County	121
IN	Perry County	123
IN	Pike County	125
IN	Porter County	127
IN	Posey County	129
IN	Pulaski County	131
IN	Putnam County	133
IN	Randolph County	135
IN	Ripley County	137
IN	Rush County	139
IN	St. Joseph County	141
IN	Scott County	143
IN	Shelby County	145
IN	Spencer County	147
IN	Starke County	149
IN	Steuben County	151
IN	Sullivan County	153
IN	Switzerland County	155
IN	Tippecanoe County	157
IN	Tipton County	159
IN	Union County	161
IN	Vanderburgh County	163
IN	Vermillion County	165
IN	Vigo County	167
IN	Wab ash County	169
IN	Warren County	171
IN	Warrick County	173
IN	Washington County	175
IN	Wayne County	177
IN	Wells County	179
IN	White County	181
IN	Whitley County	183
KS	Allen County	001
KS	Anderson County	003
KS	Atchison County	005
KS	Barber County	007
KS	Barton County	009
KS	Bourbon County	011
KS	Brown County	013
KS	Butler County	015
KS	Chase County	017
KS	Chautauqua County	019
KS	Cherokee County	021
KS	Cheyenne County	023
KS	Clark County	025
KS	Cley County	027
KS	Cloud County	029
KS	Coffey County	031
KS	Comanche County	033
KS	Cowley County	035
KS	Crawford County	037
KS	Decatur County	039
KS	Dickinson County	041
KS	Doniphan County	043
KS	Douglas County	045
KS	Edwards County	047
KS	Elk County	049
KS	Ellis County	051
KS	Ellsworth County	053
KS	Finney County	055
KS	Ford County	057
KS	Franklin County	059
KS	Geary County	061
KS	Gove County	063
KS	Graham County	065
KS	Grant County	067
KS	Gray County	069
KS	Greeley County	071
KS	Greenwood County	073
KS	Hamilton County	075
KS	Harper County	077
KS	Harvey County	079
KS	Haskell County	081
KS	Hodgeman County	083
KS	Jackson County	085
KS	Jefferson County	087
KS	Jewell County	089
KS	Johnson County	091
KS	Kearny County	093
KS	Kingman County	095
KS	Kiowa County	097
KS	Labette County	099
KS	Lane County	101
KS	Leavenworth County	103
KS	Lincoln County	105
KS	Linn County	107
KS	Logan County	109
KS	Lyon County	111
KS	McPherson County	113
KS	Marion County	115
KS	Marshall County	117
KS	Meade County	119
KS	Miami County	121
KS	Mitchell County	123
KS	Montgomery County	125
KS	Morris County	127
KS	Morton County	129
KS	Nemaha County	131
KS	Neosho County	133
KS	Ness County	135
KS	Norton County	137
KS	Osage County	139
KS	Osborne County	141
KS	Ottawa County	143
KS	Pawnee County	145
KS	Phillips County	147
KS	Pottawatomie County	149
KS	Pratt County	151
KS	Rawlins County	153
KS	Reno County	155
KS	Republic County	157
KS	Rice County	159
KS	Riley County	161
KS	Rooks County	163
KS	Rush County	165
KS	Russell County	167
KS	Saline County	169
KS	Scott County	171
KS	Sedgwick County	173
KS	Seward County	175
KS	Shawnee County	177
KS	Sheridan County	179
KS	Sherman County	181
KS	Smith County	183
KS	Stafford County	185
KS	Stanton County	187
KS	Stevens County	189
KS	Sumner County	191
KS	Thomas County	193
KS	Trego County	195
KS	Wabaunsee County	197
KS	Wallace County	199
KS	Washington County	201
KS	Wichita County	203
KS	Wilson County	205
KS	Woodson County	207
KS	Wyandotte County	209
KY	Adair County	001
KY	Allen County	003
KY	Anderson County	005
KY	Ballard County	007
KY	Barren County	009
KY	Bath County	011
KY	Bell County	013
KY	Boone County	015
KY	Bourbon County	017
KY	Boyd County	019
KY	Boyle County	021
KY	Bracken County	023
KY	Breathitt County	025
KY	Breckinridge County	027
KY	Bullitt County	029
KY	Butler County	031
KY	Caldwell County	033
KY	Calloway County	035
KY	Campbell County	037
KY	Carlisle County	039
KY	Carroll County	041
KY	Carter County	043
KY	Casey County	045
KY	Christian County	047
KY	Clark County	049
KY	Clay County	051
KY	Clinton County	053
KY	Crittenden County	055
KY	Cumberland County	057
KY	Daviess County	059
KY	Edmonson County	061
KY	Elliott County	063
KY	Estill County	065
KY	Fayette County	067
KY	Fleming County	069
KY	Floyd County	071
KY	Franklin County	073
KY	Fulton County	075
KY	Gallatin County	077
KY	Garrard County	079
KY	Grant County	081
KY	Graves County	083
KY	Grayson County	085
KY	Green County	087
KY	Greenup County	089
KY	Hancock County	091
KY	Hardin County	093
KY	Harlan County	095
KY	Harrison County	097
KY	Hart County	099
KY	Henderson County	101
KY	Henry County	103
KY	Hickman County	105
KY	Hopkins County	107
KY	Jackson County	109
KY	Jefferson County	111
KY	Jessamine County	113
KY	Johnson County	115
KY	Kenton County	117
KY	Knott County	119
KY	Knox County	121
KY	Larue County	123
KY	Laurel County	125

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
KY	Lawrence County	127
KY	Lee County	129
KY	Leslie County	131
KY	Letcher County	133
KY	Lewis County	135
KY	Lincoln County	137
KY	Livingson County	139
KY	Logan County	141
KY	Lyon County	143
KY	McCracken County	145
KY	McCreary County	147
KY	McLean County	149
KY	Madison County	151
KY	Magoffin County	153
KY	Marion County	155
KY	Marshall County	157
KY	Martin County	159
KY	Mason County	161
KY	Meade County	163
KY	Menifee County	165
KY	Mercer County	167
KY	Metcalfe County	169
KY	Monroe County	171
KY	Montgomery County	173
KY	Morgan County	175
KY	Muhlenberg County	177
KY	Nelson County	179
KY	Nicholas County	181
KY	Ohio County	183
KY	Oldham County	185
KY	Owen County	187
KY	Owsley County	189
KY	Pendleton County	191
KY	Perry County	193
KY	Pike County	195
KY	Powell County	197
KY	Pulaski County	199
KY	Robertson County	201
KY	Rockcastle County	203
KY	Rowan County	205
KY	Russell County	207
KY	Scott County	209
KY	Shelby County	211
KY	Simpson County	213
KY	Spencer County	215
KY	Taylor County	217
KY	Todd County	219
KY	Trigg County	221
KY	Trimble County	223
KY	Union County	225
KY	Warren County	227
KY	Washington County	229
KY	Wayne County	231
KY	Webster County	233
KY	Whitley County	235
KY	Wolfe County	237
KY	Woodford County	239
LA	Acadia Parish	001
LA	Allen Parish	003
LA	Ascension Parish	005
LA	Assumption Parish	007
LA	Avoyelles Parish	009
LA	Beauregard Parish	011
LA	Bienville Parish	013
LA	Bossier Parish	015
LA	Caddo Parish	017
LA	Calcasieu Parish	019
LA	Caldwell Parish	021
LA	Cameron Parish	023
LA	Catahoula Parish	025
LA	Claiborne Parish	027
LA	Concordia Parish	029
LA	De Soto Parish	031
LA	East Baton Rouge Parish	033
LA	East Carroll Parish	035
LA	East Feliciana Parish	037
LA	Evangeline Parish	039
LA	Franklin Parish	041
LA	Grant Parish	043
LA	Iberia Parish	045
LA	Iberville Parish	047
LA	Jackson Parish	049
LA	Jefferson Parish	051
LA	Jefferson Davis Parish	053
LA	Lafayette Parish	055
LA	Lafourche Parish	057
LA	La Salle Parish	059
LA	Lincoln Parish	061
LA	Livingston Parish	063
LA	Madison Parish	065
LA	Morehouse Parish	067
LA	Natchitoches Parish	069
LA	Orleans Parish	071
LA	Ouachita Parish	073
LA	Plaquemines Parish	075
LA	Pointe Coupee Parish	077
LA	Rapides Parish	079
LA	Red River Parish	081
LA	Richland Parish	083
LA	Sabine Parish	085
LA	St. Bernard Parish	087
LA	St. Charles Parish	089
LA	St. Helena Parish	091
LA	St. James Parish	093
LA	St. John the Baptist Parish	095
LA	St. Landry Parish	097
LA	St. Martin Parish	099
LA	St. Mary Parish	101
LA	St. Tammany Parish	103
LA	Tangipahoa Parish	105
LA	Tensas Parish	107
LA	Terrebonne Parish	109
LA	Union Parish	111
LA	Vermilion Parish	113
LA	Vernon Parish	115
LA	Washington Parish	117
LA	Webster Parish	119
LA	West Baton Rouge Parish	121
LA	West Carroll Parish	123
LA	West Feliciana Parish	125
LA	Winn Parish	127
MA	Barnstable County	001
MA	Berkshire County	003
MA	Bristol County	005
MA	Dukes County	007
MA	Essex County	009
MA	Franklin County	011
MA	Hampden County	013
MA	Hampshire County	015
MA	Middlesex County	017
MA	Nantucket County	019
MA	Norfolk County	021
MA	Plymouth County	023
MA	Suffolk County	025
MA	Worcester County	027
MD	Allegany County	001
MD	Anne Arundel County	003
MD	Baltimore County	005
MD	Calvert County	009
MD	Caroline County	011
MD	Carroll County	013
MD	Cecil County	015
MD	Charles County	017
MD	Dorchester County	019
MD	Frederick County	021
MD	Garrett County	023
MD	Harford County	025
MD	Howard County	027
MD	Kent County	029
MD	Montgomery County	031
MD	Prince George's County	033
MD	Queen Anne's County	035
MD	St. Mary's County	037
MD	Somerset County	039
MD	Talbot County	041
MD	Washington County	043
MD	Wicomico County	045
MD	Worcester County	047
MD	Baltimore city	510
ME	Androscoggin County	001
ME	Aroostook County	003
ME	Cumberland County	005
ME	Franklin County	007
ME	Hancock County	009
ME	Kennebec County	011
ME	Knox County	013
ME	Lincoln County	015
ME	Oxford County	017
ME	Penobscot County	019
ME	Piscataquis County	021
ME	Sagadahoc County	023
ME	Somerset County	025
ME	Waldo County	027
ME	Washington County	029
ME	York County	031
MI	Alcona County	001
MI	Alger County	003
MI	Allegan County	005
MI	Alpena County	007
MI	Antrim County	009
MI	Arenac County	011
MI	Baraga County	013
MI	Barry County	015
MI	Bay County	017
MI	Benzie County	019
MI	Berrien County	021
MI	Branch County	023
MI	Calhoun County	025
MI	Cass County	027
MI	Charlevoix County	029
MI	Cheboygan County	031
MI	Chippewa County	033
MI	Clare County	035
MI	Clinton County	037
MI	Crawford County	039
MI	Delta County	041
MI	Dickinson County	043
MI	Eaton County	045
MI	Emmet County	047
MI	Genesee County	049
MI	Gladwin County	051
MI	Gogebic County	053
MI	Grand Traverse County	055
MI	Gratio County	057
MI	Hillsdale County	059
MI	Houghton County	061
MI	Huron County	063
MI	Ingham County	065
MI	Ionia County	067
MI	Iosco County	069

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code	State	County/Area	County/ Area FIPS Code	State	County/Area	County/ Area FIPS Code
MI	Iron County	071	MN	Freeborn County	047	MO	Bates County	013
MI	Isabella County	073	MN	Goodhue County	049	MO	Benton County	015
MI	Jackson County	075	MN	Grant County	051	MO	Bollinger County	017
MI	Kalamazoo County	077	MN	Hennepin County	053	MO	Boone County	019
MI	Kalkaska County	079	MN	Houston County	055	MO	Buchanan County	021
MI	Kent County	081	MN	Hubbard County	057	MO	Butler County	023
MI	Keweenaw County	083	MN	Isanti County	059	MO	Caldwell County	025
MI	Lake County	085	MN	Itasca County	061	MO	Callaway County	027
MI	Lapeer County	087	MN	Jackson County	063	MO	Camden County	029
MI	Leelanau County	089	MN	Kanabec County	065	MO	Cape Girardeau County	031
MI	Lenawee County	091	MN	Kandiyohi County	067	MO	Carroll County	033
MI	Livingston County	093	MN	Kittson County	069	MO	Carter County	035
MI	Luce County	095	MN	Koochiching County	071	MO	Cass County	037
MI	Mackinac County	097	MN	Lac qui Parle County	073	MO	Cedar County	039
MI	Macomb County	099	MN	Lake County	075	MO	Chariton County	041
MI	Manistee County	101	MN	Lake of the Woods County	077	MO	Christian County	043
MI	Marquette County	103	MN	Le Sueur County	079	MO	Clark County	045
MI	Mason County	105	MN	Lincoln County	081	MO	Clay County	047
MI	Mecosta County	107	MN	Lyon County	083	MO	Clinton County	049
MI	Menominee County	109	MN	McLeod County	085	MO	Cole County	051
MI	Midland County	111	MN	Mahnomen County	087	MO	Cooper County	053
MI	Missaukee County	113	MN	Marshall County	089	MO	Crawford County	055
MI	Monroe County	115	MN	Martin County	091	MO	Dade County	057
MI	Montcalm County	117	MN	Meeker County	093	MO	Dallas County	059
MI	Montmorency County	119	MN	Mille Lacs County	095	MO	Daviess County	061
MI	Muskegon County	121	MN	Morrison County	097	MO	DeKalb County	063
MI	Newaygo County	123	MN	Mower County	099	MO	Dent County	065
MI	Oakland County	125	MN	Murray County	101	MO	Douglas County	067
MI	Oceana County	127	MN	Nicollet County	103	MO	Dunklin County	069
MI	Ogemaw County	129	MN	Nobles County	105	MO	Franklin County	071
MI	Ontonagon County	131	MN	Norman County	107	MO	Gasconade County	073
MI	Osceola County	133	MN	Olmsted County	109	MO	Gentry County	075
MI	Oscoda County	135	MN	Otter Tail County	111	MO	Greene County	077
MI	Otsego County	137	MN	Pennington County	113	MO	Grundy County	079
MI	Ottawa County	139	MN	Pine County	115	MO	Harrison County	081
MI	Presque Isle County	141	MN	Pipestone County	117	MO	Henry County	083
MI	Roscommon County	143	MN	Polk County	119	MO	Hickory County	085
MI	Saginaw County	145	MN	Poppe County	121	MO	Holt County	087
MI	St. Clair County	147	MN	Ramsey County	123	MO	Howard County	089
MI	St. Joseph County	149	MN	Red Lake County	125	MO	Howell County	091
MI	Sanilac County	151	MN	Redwood County	127	MO	Iron County	093
MI	Schoolcraft County	153	MN	Renville County	129	MO	Jackson County	095
MI	Shiawassee County	155	MN	Rice County	131	MO	Jasper County	097
MI	Tuscola County	157	MN	Rock County	133	MO	Jefferson County	099
MI	Van Buren County	159	MN	Roseau County	135	MO	Johnson County	101
MI	Washtenaw County	161	MN	St. Louis County	137	MO	Knox County	103
MI	Wayne County	163	MN	Scott County	139	MO	Laclede County	105
MI	Wexford County	165	MN	Sherburne County	141	MO	Lafayette County	107
MN	Aitkin County	001	MN	Sibley County	143	MO	Lawrence County	109
MN	Anoka County	003	MN	Stearns County	145	MO	Lewis County	111
MN	Becker County	005	MN	Steele County	147	MO	Lincoln County	113
MN	Beltrami County	007	MN	Stevens County	149	MO	Linn County	115
MN	Benton County	009	MN	Swift County	151	MO	Livingston County	117
MN	Big Stone County	011	MN	Tood County	153	MO	McDonald County	119
MN	Blue Earth County	013	MN	Traverse County	155	MO	Macon County	121
MN	Brown County	015	MN	Wabasha County	157	MO	Madison County	123
MN	Carlton County	017	MN	Wadena County	159	MO	Maries County	125
MN	Carver County	019	MN	Waseca County	161	MO	Marion County	127
MN	Cass County	021	MN	Washington County	163	MO	Mercer County	129
MN	Chippewa County	023	MN	Watonwan County	165	MO	Miller County	131
MN	Chisago County	025	MN	Wilkin County	167	MO	Mississippi County	133
MN	Clay County	027	MN	Winona County	169	MO	Moniteau County	135
MN	Clearwater County	029	MN	Wright County	171	MO	Monroe County	137
MN	Cook County	031	MN	Yellow Medicine County	173	MO	Montgomery County	139
MN	Cottonwood County	033	MO	Adair County	001	MO	Morgan County	141
MN	Crow Wing County	035	MO	Andrew County	003	MO	New Madrid County	143
MN	Dakota County	037	MO	Atchison County	005	MO	Newton County	145
MN	Dodge County	039	MO	Audrain County	007	MO	Nodaway County	147
MN	Douglas County	041	MO	Barry County	009	MO	Oregon County	149
MN	Faribault County	043	MO	Barton County	011	MO	Osage County	151
MN	Fillmore County	045				MO	Ozark County	153

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
MO	Pemiscot County	155
MO	Perry County	157
MO	Pettis County	159
MO	Phelps County	161
MO	Pike County	163
MO	Platte County	165
MO	Polk County	167
MO	Pulaski County	169
MO	Putnam County	171
MO	Ralls County	173
MO	Randolph County	175
MO	Ray County	177
MO	Reynolds County	179
MO	Ripley County	181
MO	St. Charles County	183
MO	St. Clair County	185
MO	Ste. Genevieve County	186
MO	St. Francois County	187
MO	St. Louis County	189
MO	Saline County	195
MO	Schuyler County	197
MO	Scotland County	199
MO	Scott County	201
MO	Shannon County	203
MO	Shelby County	205
MO	Stoddard County	207
MO	Stone County	209
MO	Sullivan County	211
MO	Taney County	213
MO	Texas County	215
MO	Vernon County	217
MO	Warren County	219
MO	Washington County	221
MO	Wayne County	223
MO	Webster County	225
MO	Worth County	227
MO	Wright County	229
MO	St. Louis city	510
MS	Adams County	001
MS	Alcorn County	003
MS	Amite County	005
MS	Attala County	007
MS	Benton County	009
MS	Bolivar County	011
MS	Calhoun County	013
MS	Carroll County	015
MS	Chickesaw County	017
MS	Choctaw County	019
MS	Claiborne County	021
MS	Clarke County	023
MS	Clay County	025
MS	Coahoma County	027
MS	Copiah County	029
MS	Covington County	031
MS	DeSoto County	033
MS	Forrest County	035
MS	Franklin County	037
MS	George County	039
MS	Greene County	041
MS	Grenada County	043
MS	Hancock County	045
MS	Harrison County	047
MS	Hinds County	049
MS	Holmes County	051
MS	Humphreys County	053
MS	Issaquena County	055
MS	Itawamba County	057
MS	Jackson County	059
MS	Jasper County	061
MS	Jefferson County	063
MS	Jefferson Davis County	065
MS	Jones County	067
MS	Kemper County	069
MS	Lafayette County	071
MS	Lamar County	073
MS	Lauderdale County	075
MS	Lawrence County	077
MS	Leake County	079
MS	Lee County	081
MS	Leflore County	083
MS	Lincoln County	085
MS	Lowndes County	087
MS	Madison County	089
MS	Marion County	091
MS	Marshall County	093
MS	Monroe County	095
MS	Montgomery County	097
MS	Neshoba County	099
MS	Newton County	101
MS	Noxubee County	103
MS	Oktibbeha County	105
MS	Panola County	107
MS	Pearl River County	109
MS	Perry County	111
MS	Pike County	113
MS	Pontotoc County	115
MS	Prentiss County	117
MS	Quitman County	119
MS	Rankin County	121
MS	Scott County	123
MS	Sharkey County	125
MS	Simpson County	127
MS	Smith County	129
MS	Stone County	131
MS	Sunflower County	133
MS	Tallahatchie County	135
MS	Tate County	137
MS	Tippah County	139
MS	Tishomingo County	141
MS	Tunica County	143
MS	Union County	145
MS	Walthall County	147
MS	Warren County	149
MS	Washington County	151
MS	Wayne County	153
MS	Webster County	155
MS	Wilkinson County	157
MS	Winston County	159
MS	Yalobusha County	161
MS	Yazoo County	163
MT	Beaverhead County	001
MT	Big Horn County	003
MT	Blaine County	005
MT	Broadwater County	007
MT	Carbon County	009
MT	Carer County	011
MT	Cascade County	013
MT	Chouteau County	015
MT	Custer County	017
MT	Daniels County	019
MT	Dawson County	021
MT	Deer Lodge County	023
MT	Fallon County	025
MT	Fergus County	027
MT	Flathead County	029
MT	Gallatin County	031
MT	Garfield County	033
MT	Glacier County	035
MT	Golden Valley County	037
MT	Granite County	039
MT	Hill County	041
MT	Jefferson County	043
MT	Judith Basin County	045
MT	Lake County	047
MT	Lewis and Clark County	049
MT	Liberty County	051
MT	Lincoln County	053
MT	McCone County	055
MT	Madison County	057
MT	Meagher County	059
MT	Mineral County	061
MT	Missoula County	063
MT	Musselshell County	065
MT	Park County	067
MT	Petroleum County	069
MT	Phillips County	071
MT	Pondera County	073
MT	Powder River County	075
MT	Powell County	077
MT	Prairie County	079
MT	Ravalli County	081
MT	Richland County	083
MT	Roosevelt County	085
MT	Rosebud County	087
MT	Sanders County	089
MT	Sheridan County	091
MT	Silver Bow County	093
MT	Stillwater County	095
MT	Sweet Grass County	097
MT	Teton County	099
MT	Toole County	101
MT	Treasure County	103
MT	Valley County	105
MT	Wheatland County	107
MT	Wibaux County	109
MT	Yellowstone County	111
MT	Yellowstone National Park	113
NC	Alamance County	001
NC	Alexander County	003
NC	Alleghany County	005
NC	Anson County	007
NC	Ashe County	009
NC	Avery County	011
NC	Beaufort County	013
NC	Bertie County	015
NC	Bladen County	017
NC	Brunswick County	019
NC	Buncombe County	021
NC	Burke County	023
NC	Cabarrus County	025
NC	Caldwell County	027
NC	Camden County	029
NC	Carteret County	031
NC	Caswell County	033
NC	Catawba County	035
NC	Chatham County	037
NC	Cherokee County	039
NC	Chowan County	041
NC	Clay County	043
NC	Cleveland County	045
NC	Columbus County	047
NC	Craven County	049
NC	Cumberland County	051
NC	Currituck County	053
NC	Dare County	055
NC	Davidson County	057
NC	Davie County	059
NC	Duplin County	061
NC	Durham County	063
NC	Edgecombe County	065
NC	Forsyth County	067
NC	Franklin County	069

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
NC	Gaston County	071
NC	Gates County	073
NC	Graham County	075
NC	Granville County	077
NC	Greene County	079
NC	Guilford County	081
NC	Halifax County	083
NC	Harnett County	085
NC	Haywood County	087
NC	Henderson County	089
NC	Hertford County	091
NC	Hoke County	093
NC	Hyde County	095
NC	Iredell County	097
NC	Jackson County	099
NC	Johnston County	101
NC	Jones County	103
NC	Lee County	105
NC	Lenoir County	107
NC	Lincoln County	109
NC	McDowell County	111
NC	Macon County	113
NC	Madison County	115
NC	Martin County	117
NC	Mecklenburg County	119
NC	Mitchell County	121
NC	Montgomery County	123
NC	Moore County	125
NC	Nash County	127
NC	New Hanover County	129
NC	Northampton County	131
NC	Onslow County	133
NC	Orange County	135
NC	Pamlico County	137
NC	Pasquotank County	139
NC	Pender County	141
NC	Perquimans County	143
NC	Person County	145
NC	Pitt County	147
NC	Polk County	149
NC	Randolph County	151
NC	Richmond County	153
NC	Robeson County	155
NC	Rockingham County	157
NC	Rowan County	159
NC	Rutherford County	161
NC	Sampson County	163
NC	Scotland County	165
NC	Stanly County	167
NC	Stokes County	169
NC	Surry County	171
NC	Swain County	173
NC	Transylvania County	175
NC	Tyrrell County	177
NC	Union County	179
NC	Vance County	181
NC	Wake County	183
NC	Warren County	185
NC	Washington County	187
NC	Watauga County	189
NC	Wayne County	191
NC	Wilkes County	193
NC	Wilson County	195
NC	Yadkin County	197
NC	Yancey County	199
ND	Adams County	001
ND	Barnes County	003
ND	Benson County	005
ND	Billings County	007
ND	Bottineau County	009
ND	Bowman County	011
ND	Burke County	013
ND	Burleigh County	015
ND	Cass County	017
ND	Cavalier County	019
ND	Dickey County	021
ND	Divide County	023
ND	Dunn County	025
ND	Eddy County	027
ND	Emmons County	029
ND	Foster County	031
ND	Golden Valley County	033
ND	Grand Forks County	035
ND	Grand County	037
ND	Griggs County	039
ND	Hettinger County	041
ND	Kidder County	043
ND	LaMoure County	045
ND	Logan County	047
ND	McHenry County	049
ND	McIntosh County	051
ND	McKenzie County	053
ND	McLean County	055
ND	Mercer County	057
ND	Morton County	059
ND	Mountrail County	061
ND	Nelson County	063
ND	Oliver County	065
ND	Pembina County	067
ND	Pierce County	069
ND	Ramsey County	071
ND	Ransom County	073
ND	Renville County	075
ND	Richland County	077
ND	Rolette County	079
ND	Sargent County	081
ND	Sheridan County	083
ND	Sioux County	085
ND	Slope County	087
ND	Stark County	089
ND	Steele County	091
ND	Stutsman County	093
ND	Towner County	095
ND	Traill County	097
ND	Walsh County	099
ND	Ward County	101
ND	Wells County	103
ND	Williams County	105
NE	Adams County	001
NE	Antelope County	003
NE	Arthur County	005
NE	Banner County	007
NE	Blaine County	009
NE	Boone County	011
NE	Box Butte County	013
NE	Boyd County	015
NE	Brown County	017
NE	Buffalo County	019
NE	Burt County	021
NE	Butler County	023
NE	Cass County	025
NE	Cedar County	027
NE	Chase County	029
NE	Cherry County	031
NE	Cheyenne County	033
NE	Clay County	035
NE	Colfax County	037
NE	Cuming County	039
NE	Custer County	041
NE	Dakota County	043
NE	Dawes County	045
NE	Dawson County	047
NE	Deuel County	049
NE	Dixon County	051
NE	Dodge County	053
NE	Douglas County	055
NE	Dundy County	057
NE	Filmore County	059
NE	Franklin County	061
NE	Frontier County	063
NE	Furnas County	065
NE	Gage County	067
NE	Garden County	069
NE	Garfield County	071
NE	Gosper County	073
NE	Grant County	075
NE	Greeley County	077
NE	Hall County	079
NE	Hamilton County	081
NE	Harlan County	083
NE	Hayes County	085
NE	Hitchcock County	087
NE	Holt County	089
NE	Hooker County	091
NE	Howard County	093
NE	Jefferson County	095
NE	Johnson County	097
NE	Kearney County	099
NE	Keith County	101
NE	Keya Paha County	103
NE	Kimball County	105
NE	Knox County	107
NE	Lancaster County	109
NE	Lincoln County	111
NE	Logan County	113
NE	Loup County	115
NE	McPherson County	117
NE	Madison County	119
NE	Merrick County	121
NE	Morill County	123
NE	Nance County	125
NE	Nemaha County	127
NE	Nuckolls County	129
NE	Otoe County	131
NE	Pawnee County	133
NE	Perkins County	135
NE	Phelps County	137
NE	Pierce County	139
NE	Platte County	141
NE	Polk County	143
NE	Red Willow County	145
NE	Richardson County	147
NE	Rock County	149
NE	Saline County	151
NE	Sarpy County	153
NE	Saunders County	155
NE	Scotts Bluff County	157
NE	Seward County	159
NE	Sheridan County	161
NE	Sherman County	163
NE	Sioux County	165
NE	Stanton County	167
NE	Thayer County	169
NE	Thomas County	171
NE	Thurston County	173
NE	Valley County	175
NE	Washington County	177
NE	Wayne County	179
NE	Webster County	181
NE	Wheeler County	183
NE	York County	185
NH	Belknap County	001
NH	Carroll County	003

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/Area FIPS Code	State	County/Area	County/Area FIPS Code	State	County/Area	County/Area FIPS Code
NH	Cheshire County	005	NV	Lyon County	019	OH	Allen County	003
NH	Coos County	007	NV	Mineral County	021	OH	Ashland County	005
NH	Grafton County	009	NV	Nye County	023	OH	Ashtabula County	007
NH	Hillsborough County	011	NV	Pershing County	027	OH	Athens County	009
NH	Merrimack County	013	NV	Storey County	029	OH	Auglaize County	011
NH	Rockingham County	015	NV	Washoe County	031	OH	Belmont County	013
NH	Strafford County	017	NV	White Pine County	033	OH	Brown County	015
NH	Sullivan County	019	NV	Carson City	510	OH	Butler County	017
NJ	Atlantic County	001	NY	Albany County	001	OH	Carroll County	019
NJ	Bergen County	003	NY	Allegany County	003	OH	Champaign County	021
NJ	Burlington County	005	NY	Bronx County	005	OH	Clark County	023
NJ	Camden County	007	NY	Broome County	007	OH	Clermont County	025
NJ	Cape May County	009	NY	Cattaraugus County	009	OH	Clinton County	027
NJ	Cumberland County	011	NY	Cayuga County	011	OH	Columbiana County	029
NJ	Essex County	013	NY	Chautauqua County	013	OH	Coshocton County	031
NJ	Gloucester County	015	NY	Chemung County	015	OH	Crawford County	033
NJ	Hudson County	017	NY	Chenango County	017	OH	Cuyahoga County	035
NJ	Hunterdon County	019	NY	Clinton County	019	OH	Darke County	037
NJ	Mercer County	021	NY	Columbia County	021	OH	Defiance County	039
NJ	Middlesex County	023	NY	Cortland County	023	OH	Delaware County	041
NJ	Monmouth County	025	NY	Delaware County	025	OH	Erie County	043
NJ	Morris County	027	NY	Dutchess County	027	OH	Fairfield County	045
NJ	Ocean County	029	NY	Erie County	029	OH	Fayette County	047
NJ	Passaic County	031	NY	Essex County	031	OH	Franklin County	049
NJ	Salem County	033	NY	Franklin County	033	OH	Fulton County	051
NJ	Somerset County	035	NY	Fulton County	035	OH	Galia County	053
NJ	Sussex County	037	NY	Genesee County	037	OH	Geauga County	055
NJ	Union County	039	NY	Greene County	039	OH	Greene County	057
NJ	Warren County	041	NY	Hamilton County	041	OH	Guernsey County	059
NM	Bernalillo County	001	NY	Herkimer County	043	OH	Hamilton County	061
NM	Catron County	003	NY	Jefferson County	045	OH	Hancock County	063
NM	Chaves County	005	NY	Kings County	047	OH	Hardin County	065
NM	Cibola County	006	NY	Lewis County	049	OH	Harrison County	067
NM	Colfax County	007	NY	Livingston County	051	OH	Henry County	069
NM	Curry County	009	NY	Madison County	053	OH	Highland County	071
NM	DeBaca County	011	NY	Monroe County	055	OH	Hocking County	073
NM	Dona Ana County	013	NY	Montgomery County	057	OH	Holmes County	075
NM	Eddy County	015	NY	Nassau County	059	OH	Huron County	077
NM	Grant County	017	NY	New York County	061	OH	Jackson County	079
NM	Guadalupe County	019	NY	Niagara County	063	OH	Jefferson County	081
NM	Harding County	021	NY	Oneida County	065	OH	Knox County	083
NM	Hidalgo County	023	NY	Onondaga County	067	OH	Lake County	085
NM	Lea County	025	NY	Ontario County	069	OH	Lawrence County	087
NM	Lincoln County	027	NY	Orange County	071	OH	Licking County	089
NM	Los Alamos County	028	NY	Orleans County	073	OH	Logan County	091
NM	Luna County	029	NY	Oswego County	075	OH	Lorain County	093
NM	McKinley County	031	NY	Otsego County	077	OH	Lucas County	095
NM	Mora County	033	NY	Putnam County	079	OH	Madison County	097
NM	Otero County	035	NY	Queens County	081	OH	Mahoning County	099
NM	Quay County	037	NY	Rensselaer County	083	OH	Marion County	101
NM	Rio Arriba County	039	NY	Richmond County	085	OH	Medina County	103
NM	Roosevelt County	041	NY	Rockland County	087	OH	Meigs County	105
NM	Sandoval County	043	NY	St. Lawrence County	089	OH	Mercer County	107
NM	San Juan County	045	NY	Saratoga County	091	OH	Miami County	109
NM	San Miguel County	047	NY	Schenectady County	093	OH	Monroe County	111
NM	Santa Fe County	049	NY	Schoharie County	095	OH	Montgomery County	113
NM	Sierra County	051	NY	Schuyler County	097	OH	Morgan County	115
NM	Socorro County	053	NY	Seneca County	099	OH	Morrow County	117
NM	Taos County	055	NY	Steuben County	101	OH	Muskingum County	119
NM	Torrance County	057	NY	Suffolk County	103	OH	Noble County	121
NM	Union County	059	NY	Sullivan County	105	OH	Ottawa County	123
NM	Valencia County	061	NY	Tioga County	107	OH	Paulding County	125
NV	Churchill County	001	NY	Tompkins County	109	OH	Perry County	127
NV	Clark County	003	NY	Ulster County	111	OH	Pickaway County	129
NV	Douglas County	005	NY	Warren County	113	OH	Pike County	131
NV	Elko County	007	NY	Washington County	115	OH	Portage County	133
NV	Esmeralda County	009	NY	Wayne County	117	OH	Preble County	135
NV	Eureka County	011	NY	Westchester County	119	OH	Putnam County	137
NV	Humboldt County	013	NY	Wyoming County	121	OH	Richland County	139
NV	Lander County	015	NY	Yates County	123	OH	Ross County	141
NV	Lincoln County	017	OH	Adams County	001	OH	Sandusky County	143

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
OH	Scioto County	145
OH	Seneca County	147
OH	Shelby County	149
OH	Stark County	151
OH	Summit County	153
OH	Trumbull County	155
OH	Tuscarawas County	157
OH	Union County	159
OH	Van Wert County	161
OH	Vinton County	163
OH	Warren County	165
OH	Washington County	167
OH	Wayne County	169
OH	William County	171
OH	Wood County	173
OH	Wyandot County	175
OK	Adair County	001
OK	Alfalfa County	003
OK	Atoka County	005
OK	Beaver County	007
OK	Beckham County	009
OK	Blaine County	011
OK	Bryan County	013
OK	Caddo County	015
OK	Canadian County	017
OK	Carter County	019
OK	Cherokee County	021
OK	Choctaw County	023
OK	Cimarron County	025
OK	Cleveland County	027
OK	Coal County	029
OK	Comanche County	031
OK	Cotton County	033
OK	Craig County	035
OK	Creek County	037
OK	Cluster County	039
OK	Delaware County	041
OK	Dewey County	043
OK	Ellis County	045
OK	Garfield County	047
OK	Garvin County	049
OK	Grady County	051
OK	Grant County	053
OK	Greer County	055
OK	Harmon County	057
OK	Harper County	059
OK	Haskell County	061
OK	Hughes County	063
OK	Jackson County	065
OK	Jefferson County	067
OK	Johnston County	069
OK	Key County	071
OK	Kingfisher County	073
OK	Kiowa County	075
OK	Latimer County	077
OK	Le Flore County	079
OK	Lincoln County	081
OK	Logan County	083
OK	Love County	085
OK	McClain County	087
OK	McCurtain County	089
OK	McIntosh County	091
OK	Major County	093
OK	Marshall County	095
OK	Mayes County	097
OK	Murray County	099
OK	Muskogee County	101
OK	Noble County	013
OK	Nowata County	105
OK	Okfuskee County	107
OK	Oklahoma County	109
OK	Okmulgee County	111
OK	Osage County	113
OK	Ottawa County	115
OK	Pawnee County	117
OK	Payne County	119
OK	Pittsburg County	121
OK	Pontotoc County	123
OK	Pottawatomie County	125
OK	Pushmataha County	127
OK	Roger Mills County	129
OK	Rogers County	131
OK	Seminole County	133
OK	Sequoyah County	135
OK	Stephens County	137
OK	Texas County	139
OK	Tillman County	141
OK	Tulsa County	143
OK	Wagoner County	145
OK	Washington County	147
OK	Washita County	149
OK	Woods County	151
OK	Woodward County	153
OR	Baker County	001
OR	Benton County	003
OR	Clackamas County	005
OR	Clatsop County	007
OR	Columbia County	009
OR	Coos County	011
OR	Crook County	013
OR	Curry County	015
OR	Deschutes County	017
OR	Douglas County	019
OR	Gilliam County	021
OR	Grant County	023
OR	Harney County	025
OR	Hood River County	027
OR	Jackson County	029
OR	Jefferson County	031
OR	Josephine County	033
OR	Klamath County	035
OR	Lake County	037
OR	Lane County	039
OR	Lincoln County	041
OR	Linn County	043
OR	Malheur County	045
OR	Marion County	047
OR	Morrow County	049
OR	Multomah County	051
OR	Polk County	053
OR	Sherman County	055
OR	Tillamook County	057
OR	Ulmatila County	059
OR	Union County	061
OR	Wallowa County	063
OR	Wasco County	065
OR	Washington County	067
OR	Wheeler County	069
OR	Yamhill County	071
PA	Adams County	001
PA	Allegheny County	003
PA	Armstrong County	005
PA	Beaver County	007
PA	Bedford County	009
PA	Berks County	011
PA	Blair County	013
PA	Bradford County	015
PA	Bucks County	017
PA	Butler County	019
PA	Cambria County	021
PA	Cameron County	023
PA	Carbon County	025
PA	Centre County	027
PA	Chester County	029
PA	Clarion County	031
PA	Clearfield County	033
PA	Clinton County	035
PA	Columbia County	037
PA	Crawford County	039
PA	Cumberland County	041
PA	Dauphin County	043
PA	Delaware County	045
PA	Elk County	047
PA	Erie County	049
PA	Fayette County	051
PA	Forest County	053
PA	Franklin County	055
PA	Fulton County	057
PA	Greene County	059
PA	Huntingdon County	061
PA	Indiana County	063
PA	Jefferson County	065
PA	Juniata County	067
PA	Lackawanna County	069
PA	Lancaster County	071
PA	Lawrence County	073
PA	Lebanon County	075
PA	Lehigh County	077
PA	Luzerne County	079
PA	Lycoming County	081
PA	Mc Kean County	083
PA	Mercer County	085
PA	Mifflin County	087
PA	Monroe County	089
PA	Montgomery County	091
PA	Montour County	093
PA	Northampton County	095
PA	Northumberland County	097
PA	Perry County	099
PA	Philadelphia County	101
PA	Pike County	103
PA	Potter County	105
PA	Schuylkill County	107
PA	Snyder County	109
PA	Somerset County	111
PA	Sullivan County	113
PA	Susquehanna County	115
PA	Tioga County	117
PA	Union County	119
PA	Venango County	121
PA	Warren County	123
PA	Washington County	125
PA	Wayne County	127
PA	Westmoreland County	129
PA	Wyoming County	131
PA	York County	133
RI	Bristol County	001
RI	Kent County	003
RI	Newport County	005
RI	Providence County	007
RI	Washington County	009
SC	Abbeville County	001
SC	Aiken County	003
SC	Allendale County	005
SC	Anderson County	007
SC	Bamberg County	009
SC	Barnwell County	011
SC	Beaufort County	013
SC	Berkeley County	015
SC	Calhoun County	017
SC	Charleston County	019
SC	Cherokee County	021
SC	Chester County	023

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/Area FIPS Code
SC	Chesterfield County	025
SC	Clarendon County	027
SC	Colleton County	029
SC	Darlington County	031
SC	Dillon County	033
SC	Dorchester County	035
SC	Edgefield County	037
SC	Fairfield County	039
SC	Florence County	041
SC	Georgetown County	043
SC	Greenville County	045
SC	Greenwood County	047
SC	Hampton County	049
SC	Horry County	051
SC	Jasper County	053
SC	Kershaw County	055
SC	Lancaster County	057
SC	Laurens County	059
SC	Lee County	061
SC	Lexington County	063
SC	McCormick County	065
SC	Marion County	067
SC	Marlboro County	069
SC	Newberry County	071
SC	Oconee County	073
SC	Orangeburg County	075
SC	Pickens County	077
SC	Richland County	079
SC	Saluda County	081
SC	Spartanburg County	083
SC	Sumter County	085
SC	Union County	087
SC	Williamsburg County	089
SC	York County	091
SD	Aurora County	003
SD	Beadle County	005
SD	Bennett County	007
SD	Bon Homme County	009
SD	Brookings County	011
SD	Brown County	013
SD	Brule County	015
SD	Buffalo County	017
SD	Butte County	019
SD	Campbell County	021
SD	Charles Mix County	023
SD	Clark County	025
SD	Clay County	027
SD	Codington County	029
SD	Corson County	031
SD	Custer County	033
SD	Davison County	035
SD	Day County	037
SD	Deuel County	039
SD	Dewey County	041
SD	Douglas County	043
SD	Edmunds County	045
SD	Fall River County	047
SD	Faulk County	049
SD	Grant County	051
SD	Gregory County	053
SD	Haakon County	055
SD	Hamlin County	057
SD	Hand County	059
SD	Hanson County	061
SD	Harding County	063
SD	Hughes County	065
SD	Hutchinson County	067
SD	Hyde County	069
SD	Jackson County	071
SD	Jerauld County	073
SD	Jones County	075
SD	Kingsbury County	077
SD	Lake County	079
SD	Lawrence County	081
SD	Lincoln County	083
SD	Lyman County	085
SD	McCook County	087
SD	McPherson County	089
SD	Marshall County	091
SD	Meade County	093
SD	Mellette County	095
SD	Miner County	097
SD	Minnehaha County	099
SD	Moody County	101
SD	Pennington County	103
SD	Perkins County	105
SD	Potter County	107
SD	Roberts County	109
SD	Sanborn County	111
SD	Shannon County	113
SD	Spink County	115
SD	Stanley County	117
SD	Sully County	119
SD	Todd County	121
SD	Tripp County	123
SD	Turner County	125
SD	Union County	127
SD	Walworth County	129
SD	Yankton County	135
SD	Ziebach County	137
TN	Anderson County	001
TN	Bedford County	003
TN	Benton County	005
TN	Bledsoe County	007
TN	Blount County	009
TN	Bradley County	011
TN	Campbell County	013
TN	Cannon County	015
TN	Carroll County	017
TN	Carter County	019
TN	Cheatham County	021
TN	Chester County	023
TN	Claiborne County	025
TN	Clay County	027
TN	Cocke County	029
TN	Coffee County	031
TN	Crockett County	033
TN	Cumberland County	035
TN	Davidson County	037
TN	Decatur County	039
TN	DeKalb County	041
TN	Dickson County	043
TN	Dyer County	045
TN	Fayette County	047
TN	Fentress County	049
TN	Franklin County	051
TN	Gibson County	053
TN	Giles County	055
TN	Grainger County	057
TN	Greene County	059
TN	Grundy County	061
TN	Hamblen County	063
TN	Hamilton County	065
TN	Hancock County	067
TN	Hardeman County	069
TN	Hardin County	071
TN	Hawkins County	073
TN	Haywood County	075
TN	Henderson County	077
TN	Henry County	079
TN	Hickman County	081
TN	Houston County	083
TN	Humphreys County	085
TN	Jackson County	087
TN	Jefferson County	089
TN	Johnson County	091
TN	Knox County	093
TN	Lake County	095
TN	Lauderdale County	097
TN	Lawrence County	099
TN	Lewis County	101
TN	Lincoln County	103
TN	Loudon County	105
TN	McMinn County	107
TN	McNairy County	109
TN	Macon County	111
TN	Madison County	113
TN	Marion County	115
TN	Marshall County	117
TN	Maury County	119
TN	Meigs County	121
TN	Monroe County	123
TN	Montgomery County	125
TN	Moore County	127
TN	Morgan County	129
TN	Obion County	131
TN	Overton County	133
TN	Perry County	135
TN	Pickett County	137
TN	Polk County	139
TN	Putnam County	141
TN	Rhea County	143
TN	Roane County	145
TN	Robertson County	147
TN	Rutherford County	149
TN	Scott County	151
TN	Sequeatchie County	153
TN	Sevier County	155
TN	Shelby County	157
TN	Smith County	159
TN	Stewart County	161
TN	Sullivan County	163
TN	Sumner County	165
TN	Tipton County	167
TN	Trousdale County	169
TN	Unicoi County	171
TN	Union County	173
TN	Van Buren County	175
TN	Warren County	177
TN	Washington County	179
TN	Wayne County	181
TN	Weakley County	183
TN	White County	185
TN	Williamson County	187
TN	Wilson County	189
TX	Anderson County	001
TX	Andrews County	003
TX	Angelina County	005
TX	Aransas County	007
TX	Archer County	009
TX	Armstrong County	011
TX	Atascosa County	013
TX	Austin County	015
TX	Bailey County	017
TX	Bandera County	019
TX	Bastrop County	021
TX	Baylor County	023
TX	Bee County	025
TX	Bell County	027
TX	Bexar County	029
TX	Blanco County	031
TX	Borden County	033
TX	Bosque County	035

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
TX	Bowie County	037
TX	Brazoria County	039
TX	Brazos County	041
TX	Brewster County	043
TX	Briscoe County	045
TX	Brooks County	047
TX	Brown County	049
TX	Burleson County	051
TX	Burnet County	053
TX	Caldwell County	055
TX	Calhoun County	057
TX	Callahan County	059
TX	Cameron County	061
TX	Camp County	063
TX	Carson County	065
TX	Cass County	067
TX	Castro County	069
TX	Chambers County	071
TX	Cherokee County	073
TX	Childress County	075
TX	Clay County	077
TX	Cochran County	079
TX	Coke County	081
TX	Coleman County	083
TX	Collin County	085
TX	Collingsworth County	087
TX	Colorado County	089
TX	Comal County	091
TX	Comanche County	093
TX	Concho County	095
TX	Cooke County	097
TX	Coryell County	099
TX	Cottle County	101
TX	Crane County	103
TX	Crockett County	105
TX	Crosby County	107
TX	Culberson County	109
TX	Dallam County	111
TX	Dallas County	113
TX	Dawson County	115
TX	Deaf Smith County	117
TX	Delta County	119
TX	Denton County	121
TX	DeWitt County	123
TX	Dickens County	125
TX	Dimmit County	127
TX	Donley County	129
TX	Duval County	131
TX	Eastland County	133
TX	Ector County	135
TX	Edwards County	137
TX	Ellis County	139
TX	El Paso County	141
TX	Erath County	143
TX	Falls County	145
TX	Fannin County	147
TX	Fayette County	149
TX	Fisher County	151
TX	Floyd County	153
TX	Foard County	155
TX	Fort Bend County	157
TX	Franklin County	159
TX	Freestone County	161
TX	Frio County	153
TX	Gainess County	165
TX	Galveston County	167
TX	Garza County	169
TX	Gillespie County	171
TX	Glasscock County	173
TX	Goliad County	175
TX	Gonzales County	177
TX	Gray County	179
TX	Grayson County	181
TX	Gregg County	183
TX	Grimes County	185
TX	Guadalupe County	187
TX	Hale County	189
TX	Hall County	191
TX	Hamilton County	193
TX	Hansford County	185
TX	Hardeman County	197
TX	Hardin County	199
TX	Harris County	201
TX	Harrison County	203
TX	Hartley County	205
TX	Haskell County	207
TX	Hays County	209
TX	Hemphill County	211
TX	Henderson County	213
TX	Hidalgo County	215
TX	Hill County	217
TX	Hockley County	219
TX	Hood County	221
TX	Hopkins County	223
TX	Houston County	;225
TX	Howard County	227
TX	Hudspeth County	229
TX	Hunt County	231
TX	Hutchinson County	233
TX	Irion County	235
TX	Jack County	237
TX	Jackson County	239
TX	Jasper County	241
TX	Jeff Davis County	243
TX	Jefferson County	245
TX	Jim Hogg County	247
TX	Jim Wells County	249
TX	Johnson County	251
TX	Jones County	253
TX	Karnes County	255
TX	Kaufman County	257
TX	Kendall County	259
TX	Kennedy County	261
TX	Kent County	263
TX	Kerr County	265
TX	Kimble County	267
TX	King County	269
TX	Kinney County	271
TX	Kleberg County	273
TX	Knox County	275
TX	Lamar County	277
TX	Lamb County	279
TX	Lampasas County	281
TX	La Salle County	283
TX	Lavaca County	285
TX	Lee County	287
TX	Leon County	289
TX	Liberty County	291
TX	Limestone County	293
TX	Limpscomb County	295
TX	Live Oak County	297
TX	Llano County	299
TX	Loving County	301
TX	Lubbock County	303
TX	Lynn County	305
TX	McCulloch County	307
TX	McLennan County	309
TX	McMullen County	311
TX	Madison County	313
TX	Marion County	315
TX	Martin County	317
TX	Mason County	319
TX	Matagorda County	321
TX	Maverick County	323
TX	Medina County	325
TX	Menard County	327
TX	Midland County	329
TX	Milam County	331
TX	Mills County	333
TX	Mitchell County	335
TX	Montague County	337
TX	Montgomery County	339
TX	Moore County	341
TX	Morris County	343
TX	Motley County	345
TX	Nacogdoches County	347
TX	Navarro County	349
TX	Newton County	351
TX	Nolan County	353
TX	Nueces County	355
TX	Ochiltree County	357
TX	Oldham County	359
TX	Orange County	361
TX	Palo Pinto County	363
TX	Panola County	365
TX	Parker County	367
TX	Parmer County	369
TX	Pecos County	371
TX	Polk County	373
TX	Potter County	375
TX	Presidio County	377
TX	Rains County	379
TX	Randall County	381
TX	Reagan County	383
TX	Real County	385
TX	Red River County	387
TX	Reeves County	389
TX	Refugio County	391
TX	Roberts County	393
TX	Robertson County	395
TX	Rockwall County	397
TX	Runnels County	399
TX	Rusk County	401
TX	Sabine County	403
TX	San Augustine County	405
TX	San Jacinto County	407
TX	San Patricio County	409
TX	San Saba County	411
TX	Schleicher County	413
TX	Scurry County	415
TX	Shackelford County	417
TX	Shelby County	419
TX	Sherman County	421
TX	Smith County	423
TX	Somervell County	425
TX	Starr County	427
TX	Stephens County	429
TX	Sterling County	431
TX	Stonewall County	433
TX	Sutton County	435
TX	Swisher County	437
TX	Tarrant County	439
TX	Taylor County	441
TX	Terrell County	443
TX	Terry County	445
TX	Throckmorton County	447
TX	Titus County	449
TX	Tom Green County	451
TX	Treviw County	453
TX	Trinity County	455
TX	Tyler County	457
TX	Upshur Parish	459
TX	Upon County	461

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
TX	Uvaide County	463
TX	Val Verde County	465
TX	Van Zandt County	467
TX	Victoria County	469
TX	Walker County	471
TX	Waller County	473
TX	Ward County	475
TX	Washington County	477
TX	Webb County	479
TX	Wharton County	481
TX	Wheeler County	483
TX	Wichita County	485
TX	Wilbarger County	487
TX	Willacy County	489
TX	Williamson County	491
TX	Wilson County	493
TX	Winkler County	495
TX	Wise County	497
TX	Wood County	499
TX	Yoakum County	501
TX	Young County	503
TX	Zapata County	505
TX	Zavala County	507
UT	Beaver County	001
UT	Box Elder County	003
UT	Cache County	005
UT	Carbon County	007
UT	Daggett County	009
UT	Davis County	011
UT	Duchesne County	013
UT	Emery County	015
UT	Garfield County	017
UT	Grand County	019
UT	Iron County	021
UT	Juab County	023
UT	Kane County	025
UT	Millard County	027
UT	Morgan County	029
UT	Piute County	031
UT	Rich County	033
UT	Salt Lake County	035
UT	San Juan County	037
UT	Sanpete County	039
UT	Sevier County	041
UT	Summit County	043
UT	Tooele County	045
UT	Uintah County	047
UT	Utah County	049
UT	Wasatch County	051
UT	Washington County	053
UT	Wayne County	055
UT	Weber County	057
VA	Accomack County	001
VA	Albemarle County	003
VA	Alleghany County	005
VA	Amelia County	007
VA	Amherst County	009
VA	Appomattox County	011
VA	Arlington County	013
VA	Augusta County	015
VA	Bath County	017
VA	Bedford County	019
VA	Bland County	021
VA	Botetourt County	023
VA	Brunswick County	025
VA	Buchanan County	027
VA	Buckingham County	029
VA	Campbell County	031
VA	Caroline County	033
VA	Carroll County	035
VA	Charles City County	036
VA	Charlotte County	037
VA	Chesterfield County	041
VA	Clarke County	043
VA	Craig County	045
VA	Culpeper County	047
VA	Cumberland County	049
VA	Dickenson County	051
VA	Dinwiddie County	053
VA	Essex County	057
VA	Fairfax County	059
VA	Fauquier County	061
VA	Floyd County	063
VA	Fluvanna County	065
VA	Franklin County	067
VA	Frederick County	069
VA	Giles County	071
VA	Gloucester County	073
VA	Goochland County	075
VA	Grayson County	077
VA	Greene County	079
VA	Greensville County	081
VA	Halifax County	083
VA	Hanover County	085
VA	Henrico County	087
VA	Henry County	089
VA	Highland County	091
VA	Isle of Wight County	093
VA	James City County	095
VA	King and Queen County	097
VA	King George County	099
VA	King William County	101
VA	Lancaster County	103
VA	Lee County	105
VA	Loudoun County	107
VA	Louisa County	109
VA	Lunenburg County	111
VA	Madison County	113
VA	Mathews County	115
VA	Mecklenburg County	117
VA	Middlesex County	119
VA	Montgomery County	121
VA	Nelson County	125
VA	New Kent County	127
VA	Northampton County	131
VA	Northumberland County	133
VA	Nottoway County	135
VA	Orange County	137
VA	Page County	139
VA	Patrick County	141
VA	Pittsylvania County	143
VA	Powhatan County	145
VA	Prince Edward County	147
VA	Prince George County	149
VA	Prince William County	153
VA	Pulaski County	155
VA	Rappahannock County	157
VA	Richmond County	159
VA	Roanoke County	161
VA	Rockbridge County	163
VA	Rockingham County	165
VA	Russell County	167
VA	Scott County	169
VA	Shenandoah County	171
VA	Smyth County	173
VA	Southampton County	175
VA	Spotsylvania County	177
VA	Stafford County	179
VA	Surry County	181
VA	Sussex County	183
VA	Tazewell County	185
VA	Warren County	187
VA	Washington County	191
VA	Westmoreland County	193
VA	Wise County	195
VA	Wythe County	197
VA	York County	199
VA	Alexandria city	510
VA	Bedford city	515
VA	Bristol city	520
VA	Buena Vista city	530
VA	Charlottesville city	540
VA	Chesapeake city	550
VA	Clifton Forge city	560
VA	Colonial Heights city	570
VA	Convington city	580
VA	Danville city	590
VA	Emporia city	595
VA	Fairfax city	600
VA	Falls Church city	610
VA	Franklin city	620
VA	Frederickburg city	630
VA	Galax city	640
VA	Hampton city	650
VA	Harrisonburg city	660
VA	Hopewell city	670
VA	Lexington city	678
VA	Lynchburg city	680
VA	Manassas city	683
VA	Manassas Park city	685
VA	Martinsville city	690
VA	Newport News city	700
VA	Norfolk city	710
VA	Norton city	720
VA	Petersburg city	730
VA	Poquoson city	735
VA	Portsmouth city	740
VA	Redford city	750
VA	Richmond city	760
VA	Roanoka city	770
VA	Salem city	775
VA	South Boston city	780
VA	Staunton city	790
VA	Suffolk city	800
VA	Virginia Beach city	810
VA	Waynesboro city	820
VA	Williamsburg city	830
VA	Winchester city	840
VT	Addison County	001
VT	Bennington County	003
VT	Caledonia County	005
VT	Chittenden County	007
VT	Essex County	009
VT	Franklin County	011
VT	Grand Isle County	013
VT	Lamoille County	015
VT	Orange County	017
VT	Orleans County	019
VT	Rutland County	021
VT	Washington County	023
VT	Windham County	025
VT	Windsor County	027
WA	Adams County	001
WA	Asotin County	003
WA	Benton County	005
WA	Chelan County	007
WA	Clallam County	009
WA	Clark County	011
WA	Columbia County	013
WA	Cowlitz County	015
WA	Douglas County	017
WA	Ferry County	019
WA	Franklin County	021

Table 1.12-3 (*Continued*)
EAS Location Codes

(b) County/Area Codes

State	County/Area	County/ Area FIPS Code
WA	Garfield County	023
WA	Grant County	025
WA	Grays Harbor County	027
WA	Island County	029
WA	Jefferson County	031
WA	King County	033
WA	Kitsap County	035
WA	Kittitas County	037
WA	Klickitat County	039
WA	Lewis County	041
WA	Lincoln County	043
WA	Mason County	045
WA	Okanogan County	047
WA	Pacific County	049
WA	Pend Oreille County	051
WA	Pierce County	053
WA	San Juan County	055
WA	Skagit County	057
WA	Skamaria County	059
WA	Snohomish County	061
WA	Spokane County	063
WA	Stevens County	065
WA	Thurston County	067
WA	Wahkiakum County	069
WA	Walla Walla County	071
WA	Whatcom County	073
WA	Whitman County	075
WA	Yakima County	077
WI	Adams County	001
WI	Ashland County	003
WI	Barron County	005
WI	Beyfield County	007
WI	Brown County	009
WI	Buffalo County	011
WI	Burnett County	013
WI	Calumet County	015
WI	Chippewa County	017
WI	Clark County	019
WI	Columbia County	021
WI	Crawford County	023
WI	Dane County	025
WI	Dodge County	027
WI	Door County	029
WI	Douglas County	031
WI	Dunn County	033
WI	Eau Claire County	035
WI	Florence County	037
WI	Fond du Lac County	039
WI	Forest County	041
WI	Grant County	043
WI	Green County	045
WI	Green Lake County	047
WI	Iowa County	049
WI	Iron County	051
WI	Jackson County	053
WI	Jefferson County	055
WI	Juneau County	057
WI	Kenosha County	059
WI	Kewaunee County	061
WI	La Crosse County	063
WI	Lafayette County	065
WI	Langlade County	067
WI	Lincoln County	069
WI	Manitowoc County	071
WI	Marathon County	073
WI	Marinette County	075
WI	Marquette County	077
WI	Menominee County	078
WI	Milwaukee County	079
WI	Monroe County	081
WI	Oconto County	083
WI	Oneida County	085
WI	Outagamie County	087
WI	Ozaukee County	089
WI	Pepin County	091
WI	Pierce County	093
WI	Polk County	095
WI	Portage County	097
WI	Price County	099
WI	Racine County	101
WI	Richland County	103
WI	Rock County	105
WI	Rusk County	107
WI	St. Croix County	109
WI	Sauk County	111
WI	Sawyer County	113
WI	Shawano County	115
WI	Sheboygan County	117
WI	Taylor County	119
WI	Trempealeau County	121
WI	Vernon County	123
WI	Vilas County	125
WI	Walworth County	127
WI	Washburn County	129
WI	Washington County	131
WI	Waukesha County	133
WI	Waupaca County	135
WI	Waushara County	137
WI	Winnebago County	139
WI	Wood County	141
WV	Barbour County	001
WV	Berkeley County	003
WV	Boone County	005
WV	Brexton County	007
WV	Brooke County	009
WV	Cabell County	011
WV	Calhoun County	013
WV	Clay County	015
WV	Doddridge County	017
WV	Fayette County	019
WV	Gilmer County	021
WV	Grant County	023
WV	Greenbier County	025
WV	Hampshire County	027
WV	Hancock County	029
WV	Hardy County	031
WV	Harrison County	033
WV	Jackson County	035
WV	Jefferson County	037
WV	Kanawha County	039
WV	Lewis County	041
WV	Lincoln County	043
WV	Logan County	045
WV	McDowell County	047
WV	Marion County	049
WV	Marshall County	051
WV	Mason County	053
WV	Mercer County	055
WV	Mineral County	057
WV	Mingo County	059
WV	Monongalia County	061
WV	Monroe County	063
WV	Morgan County	065
WV	Nicholas County	067
WV	Ohio County	069
WV	Pendleton County	071
WV	Pleasants County	073
WV	Pocahontas County	075
WV	Preston County	077
WV	Putnam County	079
WV	Raleigh County	081
WV	Randolph County	083
WV	Ritchie County	085
WV	Roane County	087
WV	Summers County	089
WV	Taylor County	091
WV	Tucker County	093
WV	Tyler County	095
WV	Upshur County	097
WV	Wayne County	099
WV	Webster County	101
WV	Wetzel County	103
WV	Wirt County	105
WV	Wood County	107
WV	Wyoming County	109
WY	Albany County	001
WY	Big Horn County	003
WY	Campbell County	005
WY	Carbon County	007
WY	Converse County	009
WY	Crook County	011
WY	Fremont County	013
WY	Goshen County	015
WY	Hot Springs County	017
WY	Johnson County	019
WY	Laramie County	021
WY	Lincoln County	023
WY	Natrona County	025
WY	Niobrara County	027
WY	Park County	029
WY	Platte County	031
WY	Sheridan County	033
WY	Sublette County	035
WY	Sweetwater County	037
WY	Teton County	039
WY	Uinta County	041
WY	Washakie County	043
WY	Weston County	045

Table 1.12-3
(c) County Subdivision Codes

Subdivision	Code
all	0
north west	1
north central	2
north east	3
west central	4
central	5
east central	6
south west	7
south central	8
south east	9

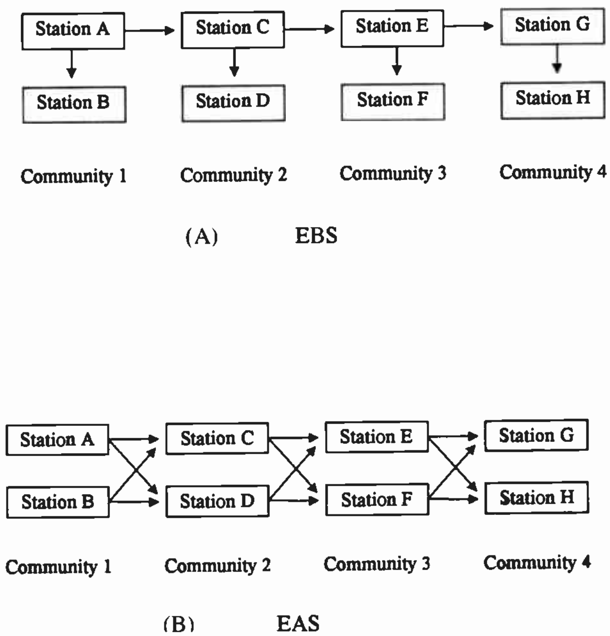


Figure 1.12-5. Comparison of emergency networks: (A) EBS system, (B) EAS system.

assigned two other stations to monitor. An example of how this design change can impact the effectiveness of the emergency network is illustrated in Figure 1.12-5.

As shown in the Figure 1.12-5 example, under the EBS there was only one path that an emergency message would take to get to Station H:

$$A \rightarrow C \rightarrow E \rightarrow G \rightarrow H$$

However, under the EAS there are eight different paths that a message takes to get to Station H:

$$A \rightarrow C \rightarrow E \rightarrow H$$
$$A \rightarrow C \rightarrow F \rightarrow H$$
$$A \rightarrow D \rightarrow E \rightarrow H$$
$$A \rightarrow D \rightarrow F \rightarrow H$$
$$B \rightarrow C \rightarrow E \rightarrow H$$
$$B \rightarrow C \rightarrow F \rightarrow H$$
$$B \rightarrow D \rightarrow E \rightarrow H$$
$$B \rightarrow D \rightarrow F \rightarrow H$$

The actual improvement in network dependability for any particular station depends on how far down the chain it was under the old EBS, and what stations it has been assigned to monitor under the new EAS. However, Figure 1.12-5, provides a general explanation of why the web structure of the new EAS is an improvement over the old system.

In the EAS, all broadcast stations fall into one of six basic categories: National Primary (NP) source, State Primary (SP) source, State Relay (SR) source, Local Primary (LP) source, Participating National (PN) source or Non-participating National (NN) source. The decisions regarding what type of source an individual station will be and what stations it must monitor are made by its State Emergency Communications Committee (SECC). Some localities may also have a Local Emergency Communications Committee (LECC) that specifies procedures for local emergencies. SECCs and LECCs are made up mostly of broadcasters who serve on a voluntary basis. These committees develop the state and local emergency communications plans for their regions. Their plans must first be approved by the FCC before they can go into effect. Once approved by the FCC a state plan, or a state plan amendment, is incorporated into the FCC's EAS mapbook for that state. The FCC mapbook for a state lists all of the broadcast stations in the state organized by their EAS local area and any special designations needed to identify each station's role in the EAS. Broadcasters who would like a copy of the FCC EAS mapbook for their state should contact the FCC's EAS Office.

The FCC Rules do give stations the option to decline to participate in the EAS. Non-participating national stations however, are still required to have EAS encoders/decoders, and they are still required to monitor two EAS sources and perform weekly and monthly tests. They must also transmit both the digital EAS codes and the two-tone attention signal during a national level alert. The only difference between a PN station and a NN station is that, during a national alert, and immediately following broadcast of the two-tone attention signal, the NN station must turn its transmitter off instead of broadcasting the actual audible alert. The NN station may only turn its transmitter back on again after the national alert is over. NN stations are not required to go off the air at all during state or local alerts-and they may participate in state or local alerts if they wish. Stations that opt to become NN sources must provide written notification of their decision to the FCC's EAS Office.

The EAS network from the perspective of a local community is illustrated in Figure 1.12-6. There are three possible entry points where an EAS alert can begin:

1. The federal government can initiate a message by feeding it to a NP source. Presently, all NP sources are broadcast stations.
2. State governments can initiate EAS messages by feeding them into SP sources. Typically, an SP source will be located in the state capital—though it can be located anywhere within the state. Most states are large enough that a single broadcaster's

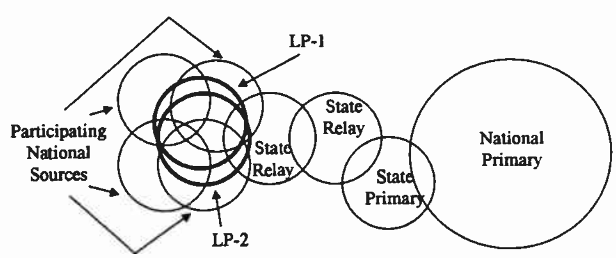


Figure 1.12-6. Typical geographic relationship of stations in the EAS network.

signal cannot cover the entire state. In these situations, it is necessary to have relay stations that pick up state level alerts and relay them out to distant counties. These relay stations are called, not surprisingly, SR sources.

3. Local governments, local emergency management offices and the National Weather Service can initiate EAS messages by feeding them into LP sources. LP sources are assigned numbers (LP-1, LP-2, etc.) to indicate the sequence in which they are to be monitored by other broadcast stations in the local area.

Individual Station Operating Requirements

The EAS responsibilities for individual stations are very simple. There are six things every station must do:

1. Own, install and keep in good working order an EAS encoder/decoder.
2. Have in its posession a copy of the *EAS Handbook* and current authentication list. (Note: at this writing the FCC and FEMA are considering eliminating the authentication list but have made no final decision).
3. Send a test message once every week, and log the fact that it was sent. The weekly test must be sent on random days within the week and at random times throughout the day or night. No weekly test is required during any week that a monthly test occurs, nor is a weekly test required during any week when a state or local emergency causes the station to transmit an EAS message. If, however, during the week the station receives one or more EAS messages concerning state or local emergencies and the station elects not to retransmit any of these messages, then the station must still broadcast the weekly EAS test message sometime during the week.
4. Log the date and time that weekly test messages, and state or local alert messages, are received from both assigned monitoring sources. If a test is not received from either one of these sources, then the reason for the failure to receive the test must be determined. This can be done by checking the station's EAS equipment, and by calling each monitoring source from which no EAS messages were received during the week. The station log must reflect the findings of any investigation into why EAS messages were not received. Also, any necessary corrective action must be taken to ensure that EAS tests, and EAS alerts, will be received properly as soon as possible.
5. Transmit the monthly test message within 15 minutes of receiving it every month, and log both the receipt and transmission of this message. The monthly test must be conducted between 8:30 am and sunset, local time, in odd-numbered months, and between sunset and 8:30 am, local time, in even-numbered months. During the week that the monthly test occurs, no weekly test is required. Monthly test messages must be rebroadcast in their entirety, except for the Transmitting Station Identifier code. Each station that rebroadcasts a monthly test message must insert its own Transmitting Station ID code. SP and LP sources originate these monthly tests. If, during a particular month, a station transmits an actual EAS emergency message that is of the same format as the monthly test, then the station does not have to transmit the monthly test.
6. Immediately broadcast any national level alerts that are received. NN stations must immediately rebroadcast the digital header and two-tone attention signal for national level alerts, followed by an announcement that indicates there is a national level alert about to be broadcast by other stations, and that this station is about to go off the air. The NN station must then go off the air for the duration of the national level alert and wait for the Emergency Action Termination message.

While not required by the FCC rules, many stations choose to voluntarily rebroadcast certain EAS messages regarding state and local emergencies. Whenever a state or local EAS message is rebroadcast, no weekly test need be transmitted that week.

Broadcasters may operate their EAS equipment in either an *automatic* or a *manual* mode. In the automatic mode the broadcaster programs the EAS unit to automatically relay certain selected EAS messages that are received by its decoder, but to only log all other received EAS messages. Stations that are operating without a transmitter duty operator will typically operate their EAS equipment in automatic mode. When EAS equipment is operated in manual mode the broadcaster must assign a person to monitor the EAS equipment and make decisions about which alerts to relay. Stations that operate their EAS equipment in manual mode are allowed to control their EAS equipment from a remote location, provided their remote control system enables them to put national level alerts on the air immediately and required monthly tests on the air within 15 minutes.

Although broadcasters can choose to air, or not to air, any state or local alert, they have no discretion when a national level alert is received. Both automatically operated and manually operated stations must interrupt their programming immediately when a national level alert is received. Stations that are participating in the EAS must put the national alert on the air. Stations that have opted not to participate in the EAS (by providing written notification to the FCC's EAS Office) must broadcast the digital EAS header codes and a short announcement then shut their transmitters off for the duration of the alert.

Every EAS message that is received by a broadcaster must be logged, regardless of whether the broadcaster decides to relay the message to its audience, and regardless of whether the station is operating in automatic or manual mode.

EAS Transmissions Via RDS

EAS information may be transmitted via an FM station's Radio Data System (RDS) subcarrier—

though transmitting EAS information via RDS does not relieve a station of its responsibility to transmit EAS information in its main audio channel. One of the advantages of using the RDS subcarrier for EAS transmissions is that it enables *silent* emergency warning networks to be developed that can provide very locally specific alerts to audiences that have RDS EAS decoding equipment. For example, using RDS a broadcaster could send out an alert about a chemical spill at a factory that would only be heard by people who are living and working in the immediate vicinity of the factory. This would be done by coding the RDS EAS message with an access code that would prevent receivers that are not programmed with the same access code from decoding the alert message.

RDS uses a 1,187.5 baud data stream that is transmitted on an FM station's 57 kHz subcarrier. The format of RDS transmissions in the United States is specified in the *United States Radio Broadcast Data System (RBDS)* Standard, which was adopted by the National Radio Systems Committee. (See Chapter 4.6, *Radio Broadcast Data System (RBDS)*, for a more complete discussion of the format of the RDS signal). The following is a description of a typical sequence of events that occurs when an EAS message is transmitted via RDS:

1. The Warning Activation bit in Group 3A of the RDS signal is set to 1 to tell sleeping battery operated receivers to wake up.
2. The Program Type code (ROCK, COUNTRY, JAZZ, etc.) is set to "ALERT!" to tell consumer RDS equipment to interrupt current activity (CD/cassette playback, radio off state, etc.) and switch to FM reception.
3. The actual audio alert message is transmitted in the FM station's main audio channel, and a text version of the alert message is transmitted to radiotext equipped RDS receivers using Groups 2A or 2B.
4. The SAME codes are transmitted using Group 9A. These codes may be useful to specialized RDS receivers equipped to decode SAME data. However, consumer RDS radios generally do not make use of this data because of the additional signal processing circuitry that would have to be included to decode the information. Consumer radios rely on the actual audio alert message, and any radiotext information that is transmitted.
5. At the end of the EAS alert, the Program Type code is set back to its appropriate state (ROCK, COUNTRY, JAZZ, etc.), and the Warning Activation bit is set back to zero to tell battery operated receivers that they may return to their *sleep* state.

CONCLUSION

For most broadcasters, the EAS is an improvement over the EBS because it enables them to automate the emergency warning function at their stations, thus making it easier for them to operate unattended transmitters and thereby reduce their operating expenses. Broadcasters that make use of RDS technology to enhance their EAS alerts can provide added services to their audiences by enabling RDS equipped radios to do such things as automatically turn on in the middle of the night to warn of approaching danger.

Section 2: Broadcast Towers and Systems

Introduction by Louis Libin, Broad Comm, Inc. Woodmere, New York

In this section, readers will be given an overview of the broadcast signal, as it begins to travel through the coaxial transmission line and takes on propagational characteristics. This section also discusses all aspects of the physical towers, design, erection and maintenance, as well as the required lightening protection for the tower structure. This section is of vital importance to the broadcast engineer because the transmission path is critical to the transmission and reception of a high quality broadcast signal. As broadcasters begin the transition from analog to digital, the highest quality of the broadcast signal will be needed from the beginning of the transmission path.

The many changes that are taking place in the broadcast industry do not mean that we can put aside the physical transmission. The physics of radio frequency propagation remain the same for digital broadcasting. The techniques relating to the transmission and reception, as well as the frequency bands will be different. The *Radio Wave Propagation* chapter begins with a valuable tutorial. For those engineers without an engineering background, the chapter will provide the framework for a real understanding of the physical concepts that play a role in wave propagation. Basic mathematical equations are included for educational purposes, and the variations in propagation, through a description of the electromagnetic spectrum are described in detail. The chapter will then explain propagation in free space and apply it to real broadcasting—just in time to usher in a better understanding of the transmission of digital technology. Broadcast engineers need more than just a cursory knowledge of the science of propagation, because many stations are now switching to the UHF band. Broadcast engineers will find that this chapter will assist them in responding to their newest audiences, the digital audiences, whose signals are now taking different routes, over the same terrain to reach the home antenna.

The transition to digital is now beginning. For many stations, the tower issues are now more of a focus than ever before. Every station needs to either replace a tower or add an antenna. The *Design, Erection and Maintenance of Antenna Structures* chapter by Thomas Hoenninger is extremely timely to broadcast engineers. This chapter provides necessary overview to familiarize the broadcast engineer with the various types of physical tower configurations, and various materials used in their construction. There are also regulatory requirements imposed on the owners of towers. This chapter focuses the broadcast engineer on the various requirements for lighting, modification, maintenance, inspections and reporting. In today's world, where the bottom line is cost, this overview is of great importance to the engineer and owner.

Broadcasters are investing more and more into their transmission facilities, including towers, waveguide and antennas. The *Lightening Protection For Tower Structures* chapter by Ed Lobnitz provides a very practical look at lightening and how the broadcaster can prepare best for this natural phenomena. Broadcasters need to protect their most valuable assets, their tower, transmission line and of course, the antenna itself from damage. The smartest design dictates that the best strategy would prevent damage from occurring; on the flip side, few crews are available, the process is expensive and towers are becoming more crowded and complex systems. The chapter is written in tutorial form, and is most significant to broadcasters. The reader will learn everything from thunderstorm cell development, and associated lightening discharge, to the best structural protection strategies available today.

In its simplest form, the transmission line guides the energy along, from one point to another. The transmission line is one of the most significant pieces of hardware in the broadcast transmission plant. It must operate with extremely high efficiencies and losses must be kept to an absolute minimum. The *Coaxial Transmission Lines* chapter by Kerry Cozad discusses the major selection criteria used for transmission line, both electrical and physical. The broadcasters who are now mandated by the Federal Communications Commission to begin to broadcast the digital signal will find this chapter extremely helpful. The chapter provides an overview on design considerations including, but not limited to, power handling and attenuation and VSWR. Physical connections, installation and testing are also described.

2.1
RADIO WAVE PROPAGATION

MARTIN H. BARRINGER
ADDITIONAL MATERIAL PROVIDED BY
KENNETH D. SPRINGER

INTRODUCTION

Radio wave propagation is the study of the transfer of energy at radio frequencies from one point, a transmitter, to another, a receiver. Radio waves are part of the broad electromagnetic spectrum that extends from the very low frequencies which are produced by electric power facilities up to the extremely high frequencies of cosmic rays. Between these two extremes are bands of frequencies that are found in every day uses: audio frequencies used in systems for the reproduction of audible sounds, radio frequencies, infrared light and ultraviolet light and x-rays.

All electromagnetic waves propagate at the same velocity, regardless of the frequency. Light is an electromagnetic wave, and thus the propagation velocity is often referred to as "the speed of light" (c), which for a vacuum is approximately 3×10^8 m/sec. The velocity of any wave is dependent upon the medium in which it is travelling, but for simplicity is usually considered with respect to a vacuum. The frequency of a wave is defined in terms of the number of cycles per second or hertz (Hz) and is related to the wavelength (λ) by the expression $f = c/\lambda$. Figure 2.1-1 shows the ranges of various bands within the electromagnetic spectrum in terms of frequency and wavelength.

Radio frequencies are generally considered to be that portion of the electromagnetic spectrum below the infrared frequencies. At present, the practical upper limit of radio frequencies is roughly 100 GHz.[1] Within the radio frequency spectrum are bands of frequencies that have been allocated to the broadcast service. The following discussions and methods will apply particularly to these bands of the radio frequency spectrum.

The AM band frequency allotments are contained in what is referred to as medium frequencies (MF), 300 kHz to 3 MHz. The FM band frequencies and a portion of the TV band are contained in the VHF band which extends from 30 MHz to 300 MHz. The remaining TV allocations are contained in the UHF band of 300 MHz to 3 GHz. Allocations for broadcast auxiliary services such as remote pickup, studio/transmitter links, intercity relays, MDS, and ITFS are interspersed within the MF, VHF, UHF and SHF (super high frequency) bands. Table 2.1-1 illustrates some of the allotments assigned to the broadcast service. The allocations for auxiliary services may change from time to time as the needs of various services for radio frequencies change and as technology for equipment improves.

QUANTIFYING PROPAGATION

The energy that is emitted from a transmitter may take many different paths before it is received. The path that the radio wave will take depends on many factors, some of which include: frequency, antenna type and height, atmospheric conditions and terrain. Radio waves that propagate along the surface of the earth are commonly referred to as ground waves. All radio waves have some ground wave component, however, because the earth is a lossy medium, it severely attenu-

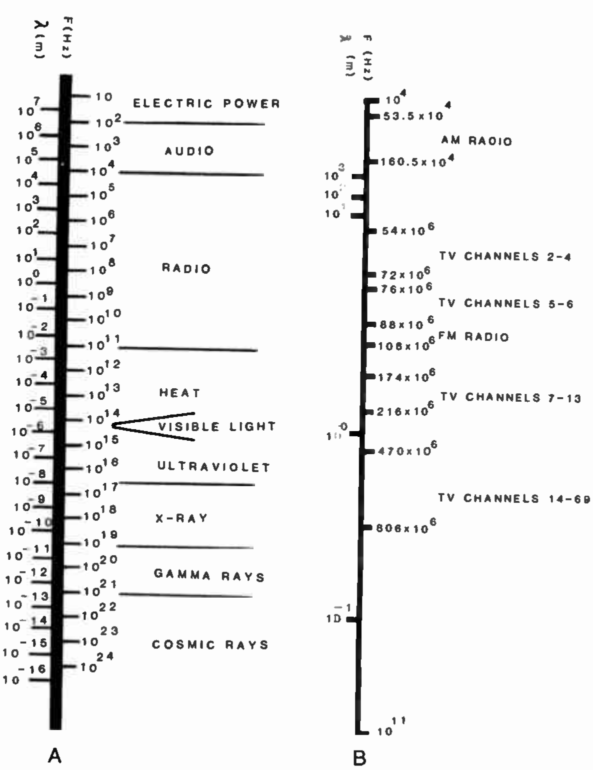


Figure 2.1-1. Electromagnetic spectrum and broadcast radio spectrum.

Table 2.1-1

Broadcast Frequency Allocations

MW 300 kHz–3 MHz	
AM:	525 kHz–1705 kHz
VHF 30 MHz–300 MHz	
FM:	88 MHz–108 MHz
TV:	54 MHz–72 MHz Channels 2–4
	76 MHz–88 MHz Channels 5–6
	174 MHz–216 MHz Channels 7–13
UHF 300 MHz–3 GHz	
TV:	470 MHz–806 MHz Channels 14–69
AM-FM STL:	947 MHz–952 MHz
MDS:	2150 MHz–2162 MHz
ITFS:	2500 MHz–2686 MHz
Auxiliary Services:	2000 MHz–3000 MHz
SHF 3 GHz–30 GHz	
Auxiliary Services:	6.425 GHz–7.125 GHz
CARS:	12.700 GHz–13.250 GHz
TV STL:	17.700 GHz–19.700 GHz

ates the radio wave. This attenuation increases with frequency, so this mode of propagation is useful only for frequencies below 30 MHz. To achieve significant distances, the atmosphere is preferred over the ground as a transmission medium. The atmosphere is comprised of several different layers, as depicted in Figure 2.1-2. The troposphere is the layer that extends from the earth's surface up to about 16 km. This layer is the chief mode of propagation for frequencies above about 30 MHz, and propagation through this layer is dependent upon weather conditions. The next layer is the stratosphere which extends to about 40 km above the earth. This layer has no major effect on the propagation of radio waves. The ionosphere extends upwards of 400 km above the surface of the earth. This region is a charged environment where the air is sufficiently ionized, mainly by the sun's ultraviolet radiation, to reflect or absorb radio waves below about 30 MHz. The ionosphere is constantly changing and is usually considered as consisting of the following sublayers.[2]

- D layer—This layer exists at heights from about 50 km to 90 km and is present only during daylight hours. The electron density is directly related to the elevation angle of the sun. This layer absorbs medium and high frequency waves.
- E layer—This layer exists at a height of about 110 km and is important in the nighttime propagation of medium frequency waves. The ionization of this layer is closely related to the elevation angle of the sun. At certain times irregular cloud-like areas of high ionization may occur. These areas are known as sporadic E and occasionally prevent frequencies that normally penetrate the E layer from reaching higher layers. The sporadic E layer is prevalent during the summer and winter months. The sporadic E layer formed during the summer is the longest lasting from May to August, and the winter layer lasts about half as long beginning in December. During the midsummer months when the electron density is at its greatest levels, TV signals in the lower VHF band can be transmitted over distances of hundred or thousands of kilometers.[4]
- F1 layer—This layer exists at heights of about 175 to 200 km and is present only during the day. Waves that usually penetrate the E layer (3 to 30 MHz) will penetrate this layer and be reflected by the F2 layer. This layer introduces additional absorption of these waves.
- F2 layer—This layer exists at the upper boundaries of the atmosphere, 250 km to 400 km, and is present at all times; though the height and electron density will vary from day to night, with the seasons, and over sunspot cycles. During the night the F1 layer merges with the F2 layer at about 300 km. This, in addition to the reduction of the D and E layers, causes nighttime field intensities and noise to be generally higher than during the day.

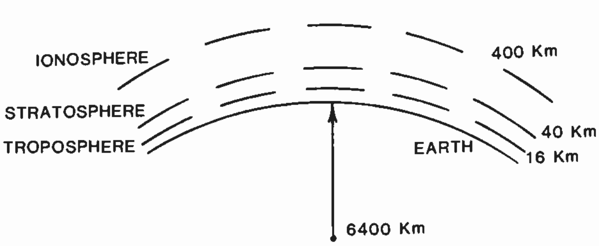


Figure 2.1-2. Atmosphere layers.

Free Space Propagation

To evaluate and compare radio wave propagation under various conditions, it is convenient to establish a reference standard. It is customary to consider as a standard the theoretically calculated loss for waves propagated in free space between two idealized antennas. The simplest case to investigate is the radiation emitted from an isotropic source: an ideal antenna which radiates energy with uniform intensity in all directions. An analogy to an isotropic antenna is a point source of light, such as a candle. The intensity of the energy varies proportionally to the inverse of the distance squared from the source, the *inverse square law*. The power of flux per unit area P_a(W/m^2) at a distance d(m) from a loss free isotropic antenna radiating a power P_1(W) is given by:

$$P_a = P_t/4\pi d^2 \qquad [1]$$

where $4\pi d^2$ is the surface area of a sphere at a distance d(m) from the source. The power available from a loss free antenna P_r is the product of the power flux per unit area (P_a), and the effective aperture area of the receiving antenna (A_e). This area is related to the gain of the antenna by the expression:

$$A_e = G\lambda^2/4\pi \qquad [2]$$

Aperture areas and gains for a specific antenna can be found in the sixth edition of the NAB *Engineering Handbook*, page 121.[3] For a loss free isotropic antenna, $G = 1$, the basic free space transmission loss is defined as:

$$L_{bf} = P_t/P_r = (4\pi d/\lambda)^2 \qquad [3]$$

where d and λ have the same units. This equation can be rewritten in its more common form, expressing the loss in dB, as:

$$L_{bf} = 32.44 + 20\log(F) + 20\log(d) \qquad [4]$$

where F is the frequency in megahertz (MHz) and d is the distance between the antennas in kilometers. In the above equation is should be remembered that ideal loss-free isotropic antennas are considered. In real world systems, antenna gain is a significant factor. The transmission loss, L, incorporates the antenna gains and is defined as:

$$L = L_{bf} - (G_t + G_r + L_d) \qquad [5]$$

Where G_t and G_r are the free space antenna gains with respect to isotropic for the transmitting and receiving antenna respectively. The term L_d is the aperture-to-medium coupling loss or polarization coupling loss between the antennas. The term L_d will have a value of 0 dB when the transmitting and receiving antenna have the same polarization.

In considering the potential service area coverage for a broadcast station, it is usually more desirable to express measurements in terms of field strength rather than transmission loss as previously presented. The root mean square (RMS) field strength, E(V/m), at a point where the power density of a plane wave is P_a(W/m^2) is given by:

$$E = \sqrt{120\pi P_a} \qquad [6]$$

where the term 120π is the impedance of free space. The field strength is related to the power available from a loss free isotropic antenna by combining Equations [1], [3] and [6] above as:

$$E = \sqrt{480\pi^2 P_r/\lambda^2} \qquad [7]$$

A more useful form of the free-space field can be expressed in logarithmic terms above 1 microvolt per meter (dBu) when F is in megahertz and P_r is expressed in decibels above 1 kW (dBK):

$$E(\text{dBu}) = 107.2 + P_r + 20\log(F)\text{dBu} \qquad [8]$$

The electric field produced by a transmitter radiating a power P_t(W) at a distance d(m) in free space can be derived from Equations [1], [3] and [6] and is given by:

$$E = \sqrt{30 P_t/d^2} \qquad [9]$$

or, in logarithmic terms, where P_t is expressed in decibels above 1 kW (dBK), d is in kilometers, and a transmitting antenna has a gain G_t in decibels above isotropic:

$$E(\text{dBu}) = 105 + P_t + G_t - 20\log(d) \qquad [10]$$

Using the same units, the field strength E(dBu) for nonfree-space environments can be related to the basic transmission loss by:

$$L_b(\text{dB}) = 137 + 20\log(F) + P_t + G_t - E \qquad [11]$$

These equations form the basis for characterizing propagation. They do not, however take into account such real world factors as the presence of the earth, atmosphere or obstructions. To adequately describe an actual radio system, additional losses will need to be added to the free-space equations derived above.

Presence of Earth

When the transmitting and receiving antennas are placed over ground, the propagation of radio waves is modified from the free-space models presented above. Radio waves that strike the earth are partially absorbed and partially reflected. Waves that are reflected by the earth experience changes in the phase of the wave, which affects the distribution of available energy. The extent to which the waves are reflected or absorbed is dependent upon frequency and the ground constants: conductivity and permittivity.

Propagation Over Plane Earth

The geometry of the idealized situation of propagation between two antennas placed above a plane earth is shown in Figure 2.1-3. This geometry is valid for antennas that are sufficiently closely located so that the curvature of the earth is not a factor, yet far enough apart from each other so that the energy may be described as a plane wave, and ray theory can be applied. The resultant received electric field can be represented as the sum of the direct and reflected rays:

$$E = E_d[1 + |R|e^{j(\phi_\Delta + \phi_r)}] \qquad [12]$$

This equation is valid for small angels of θ and deserves some additional explanation. The term E_d is the free-space electric field that is produced at a distance d(m) by the direct ray. The terms $[R]$ and ϕ_r are the magnitude and phase of the complex reflection coefficient. This term is dependent upon the nature of the surface (conductivity (δ) and permittivity (E_r)), the angle between the surface and incident wave, the wavelength of the radio wave, the polarization of the wave and the curvature of the earth. The magnitude of the reflection coefficient varies between -1 and $+1$. Several sources have derived equations for the reflection coefficient and plotted the effects of changing variables, and the reader is referred to these for further study.[4,5] The term ϕ_Δ is the phase delay due

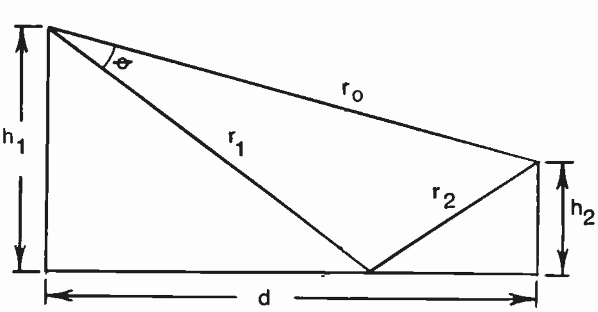


Figure 2.1-3. Path rays for plane earth.

to the longer path that must be taken by the reflected wave, and has the form of:

$$\phi_\Delta = 4\pi h_1 h_2/\lambda d \qquad [13]$$

It is often sufficient to assume the ground approximates a large flat surface. In such a case, a sufficiently accurate expression is given by:[4]

$$E = 2E_d \sin(2\pi h_1 h_2/\lambda d) \qquad [14]$$

Some cases of special merit that can be derived from Equation [14] are:

Case I	$h_1h_2 = d\lambda/2$	$E = 0$
Case II	$h_1h_2 = d\lambda/4$	$E = 2E_d$
Case III	$h_1h_2 = d\lambda/12$	$E = E_d$

Therefore, depending on the antenna heights, distances, and wavelength, it is possible to totally cancel out the field at the receiver or magnify the wave to a field strength double that which could be achieved from a free-space field. The variation of signal strength due to multipath effects can be minimized in point-to-point applications through the use of antennas with narrow beamwidths.

When considering the case of VHF antennas that are close to the ground, the effective antenna heights h_t(m) and h_r(m) will need to be substituted for h_1 and h_2 respectively for Equation (14). The new antenna heights *ht* and *hr* allow for the effects caused by the relative permittivity (ϵ_r), and conductivity (δ) of the ground. The effective antenna heights are related to the physical antenna heights above ground level by:[4]

$$h_t = \sqrt{h_1^2 + h_0^2} \qquad [15.1]$$

$$h_r = \sqrt{h_2^2 + h_0^2} \qquad [15.2]$$

the term h_0 is dependent upon the type of polarization being considered.

Vertical Polarization

$$h_0 = (\lambda/2\pi)[(\epsilon_r + 1)^2 + (60\lambda\sigma^2)]^{1/4} \qquad [16.1]$$

Horizontal Polarization

$$h_0 = (\lambda/2\pi)[(\epsilon_r - 1)^2 + (60\lambda\sigma^2)]^{-1} \qquad [16.2]$$

Table 2.1-2 lists values for conductivity and permittivity for various soil conditions. As a way of example, assume that an antenna is placed 3 m (9.8 ft) above dry, sandy, flat coastal land ($\delta = 8 \times 10^{-3}$ S/m, $\epsilon_r = 10$) and operates at a frequency of 100 MHz ($\lambda = 3$ m). Then h_0 in [16.1] and [16.2] will be 1.59 m and 0.16 m respectively. The effective height of the antenna will then be increased to 3.4 m (11.1 ft) for vertical polarization and will remain unchanged at 3 m (9.8 ft) for horizontal polarization. As the frequency increases above VHF, the wavelength becomes increasingly small and the distinction between true antenna height and effective height is immaterial.

Medium Frequency Propagation

As stated earlier, medium frequency waves lie in the frequency range of 300 kHz to 3 MHz and are characterized by their long wavelengths (1,000 to 100 meters). The standard AM broadcasting band frequencies are within this range. For the AM service, the transmitting antenna is located right at the surface of the earth, and the receiving antenna is very close to the earth's surface with respect to a wavelength. In this case the direct and ground reflected waves cancel, and the transmission is by means of the ground wave (also known as the surface wave) and the sky wave.

Table 2.1-2
Ground Conductivity and Dielectric Constants

Terrain	Conductivity, δ (S/m)	Relative Dielectric Constant, ϵ_r (esu)
Sea Water	5	80
Fresh Water	8×10^{-3}	80
Dry Sandy, flat coastal land	8×10^{-3}	10
Marshy, forested flat land	8×10^{-3}	12
Rich agricultural land, low hills	1×10^{-2}	15
Pasture land, medium hills and forest	5×10^{-3}	13
Rocky land, steep hills	2×10^{-3}	10
Mountainous	1×10^{-3}	5
Residential Area	2×10^{-3}	5
Industrial Area	1×10^{-4}	3

Ground Waves

These waves are characterized by the fact that they are guided along the earth's surface, similar to a transmission line. The field is attenuated in this propagation mode by losses in the ground. Therefore the composition of the soil, ϵ_r and δ, have a direct bearing on the amount of attenuation the wave will experience, and subsequently how far reliable communications can be established. The attenuation is also dependent upon the frequency and polarization type. The attenuation factor, A, is a measure of the amount of attenuation present and can be determined for a ground wave using the chart of Figure 2.1-4. The term ρ is known as the numerical distance and b is the phase constant. Values for these terms can be calculated from the following equations:[5]

$$\rho = (\pi d/\lambda x)\cos(b) \qquad [17.1]$$

$$b = \arctan[(\epsilon_r + 1)/x] \qquad [17.2]$$

$$x = 18 \times 10^3\sigma/F \qquad [17.3]$$

To determine the electric field strength, the attenuation factor must be added to Equation [12].

$$E = E_d[1 + re^{j(\phi_\Delta + \phi_\Delta)} + (1 - R)Ae^{j(\phi_\Delta + \phi_\Delta)}] \qquad [18]$$

It is interesting to note that the same earth which acts as a conductor at very low frequencies will act as a small-loss dielectric at very high frequencies. It is also noteworthy to observe that the losses for horizontally polarized waves are much greater than for vertically polarized waves. Thus, for practical applications only, vertically polarized waves should be considered. For more detailed and accurate representations of the ef-

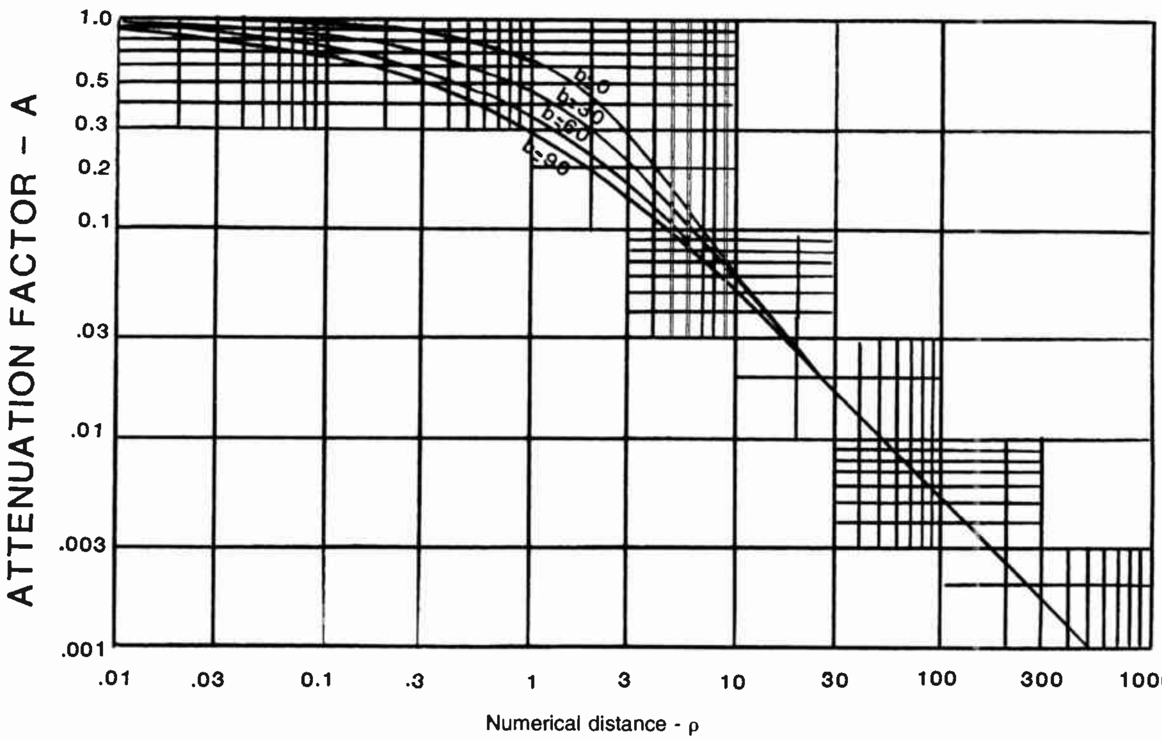


Figure 2.1-4. Attenuation factor of ground waves.

fects of ground wave, the works of Norton and as expanded upon by Jordan should be consulted.[6,5]

Sky Waves

While the ground wave provides the major path for medium frequency propagation, the wave attenuates relatively quickly with distance and is reliable for distances of only a few hundred kilometers. To achieve greater distances, the waves propagate via the ionosphere and are known as sky waves, and can provide sufficient signal strength at distances up to a few thousand kilometers.

The ionosphere is a constantly changing environment that begins approximately 65 km (40 miles) above the earth and extends to about 400 km (250 miles). This region of the atmosphere is composed of three major sublayers, D, E, and F. These layers are not present at all times. For example, the D layer is present only during the day and is a major absorber of medium-frequency waves. The E layer is a principal reflector of medium-frequency waves. Thus during the day the majority of the medium-frequency waves are absorbed by the D layer, but at night the D layer is not present, allowing the medium-frequency waves to be reflected by the E layer.

Interference Between Ground Waves and Sky Waves

Interference to a receiver may occur from co-channel stations located many kilometers from the desired station. Because of the sky wave, sufficient signal strength may be received to interfere with the local station. This effect has been minimized by the FCC by limiting two factors in the operation of some AM stations: the operating power and time of operation.

Multipath interference occurs when the waves from a transmitting antenna reach a receiver from different paths in such a manner as to cancel or severely interfere with each other. This can happen at distances where both the ground wave and sky wave are sufficiently strong to interact. The geometry of this is similar to that shown in Figure 2.1-3, except the direct ray will be a result of the ground wave and the reflected wave will be from the ionosphere. At distances relatively close to the transmitter, the ionosphere will not reflect waves back to the earth, so the ground wave is predominant. At distances beyond a few hundred kilometers, the sky wave will dominate and the ground wave will be too weak to interfere. Multipath interference can also occur where the sky wave follows more than one path to the receiver.

Effects of Solar Activity

Interference to medium-frequency waves can also be caused by solar activity such as sunspots and flares which manifest an increased or reduced emission of radiation from the sun. The changes in solar radiation levels can cause changes in the ionospheric layers that may result in unusual sky wave propagation conditions called *skip*, which, in turn, can cause inter-station interference. The effects of such solar activity will have their strongest effect on propagation in the AM band during the first 5 to 10 days after the start of a storm. This has the effect of reducing sky-wave field strengths. The effect has been observed to increase with frequency.[13]

Propagation Above 3 MHz

At frequencies above about 30 MHz the principal propagation mode is tropospheric. The surface wave is attenuated too severely to be of any practical long distance use and, though attenuated, the sky wave is usually passed through the ionosphere to space.

For waves that propagate close to the earth's surface the curvature of the earth will introduce additional effects that must be included in the plane earth model that was considered earlier Figure 2.1-5 shows the geometry of a smooth earth model. First the reflection coefficient R of the reflected wave has different characteristics that for a plane surface. Since the wave is reflected against a curved earth the energy diverges more than is predicted by the inverse square law and the reflection coefficient R, in Equation [12], must be multiplied by the divergence factor D, given by:[4]

$$D = \sqrt{1 + 2d_1d_2/2a_e(h_t' + h_r')} \qquad [19]$$

It should be noted that for smooth earth conditions, the heights h_t' and h_r', for the transmitting and receiving antennas above the plane tangent to the earth at the point of reflection are less than the antenna heights h_t and h_r above the surface of the earth.

Under normal propagation conditions, the refractive index of the atmosphere decreases with height so that radio waves near the surface of the earth travel more slowly than at high altitudes. This variation in velocity as a function of height results in a bending of the radio waves. This may be represented as a modified earth radius commonly known as the effective earth radius, a_e, which allows the radio waves to be represented as straight lines. The ratio of the effective earth radius to true earth radius is commonly known as the k factor. Values of k can vary between from 0.6 to 5.0 depending

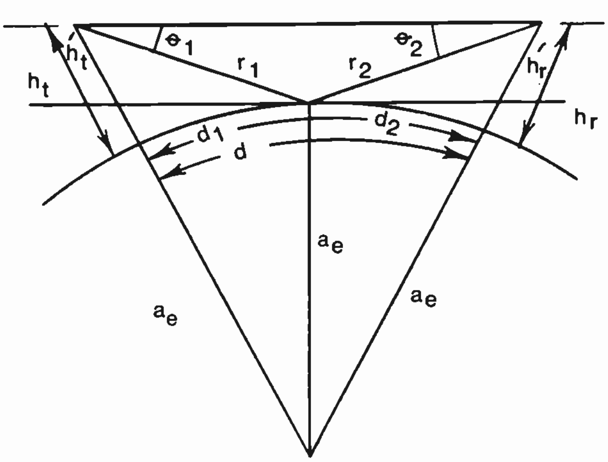


Figure 2.1-5. Reflection from smooth earth on line-of-sight path.

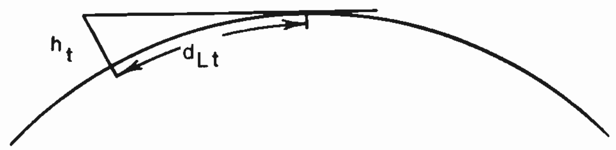


Figure 2.1-6. Distance to radio horizon.

on the climate being considered. For temperate climates the average value of k is 1.33, and most works refer to this as the 4/3 earth model when used in calculation.[2]

Beyond Line of Sight Conditions

In order to determine when conditions exist where propagation is considered to be beyond line of sight, the respective distances from the transmitter and receiver to the radio horizon must be calculated. The radio horizon is the distance the horizon appears from an antenna, as defined by a plane from the antenna to the tangent of the earth's surface and is depicted in Figure 2.1-6. The equation for the radio horizon in terms of d_{lt} (km) and h_t (m) and the k factor is of the form:

$$d_{lt} = 3.57\sqrt{h_t k} \qquad [20]$$

When the sum of the distances to the radio horizon for the transmitter and receiver is less than the total distance of the path under consideration, then a beyond line of sight condition exists. Diffraction makes it possible for radio waves to travel beyond that possible for line of sight transmission, though an additional loss term must be added to the free-space loss. The amount of attenuation can be determined by diffraction methods. The geometry of beyond line of sight propagation is shown in Figure 2.1-7.

The exact calculation of the field strength at any point beyond the line of sight for a smooth earth is rather complex and the presentation of such a method is beyond the scope of this text. However, nomograms have been developed that apply to a large number of cases. For the reader interested in the actual prediction of the losses to be expected for smooth earth diffraction, the National Bureau of Standards publication may be consulted.[9]

Figure 2.1-8 is a nomogram that can be used to determine the loss that must be added to the free-space loss. In order to use the nomogram, the distance d_{lt} must be less than d_{lr}. The total loss (L) is the sum of the three losses L1, L2, and L3. By way of example, assume a system that has the following parameters:

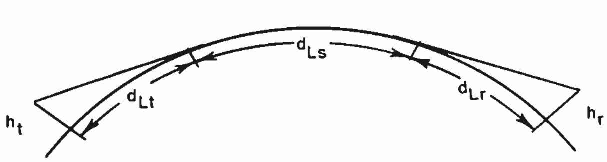


Figure 2.1-7. Beyond line of sight.

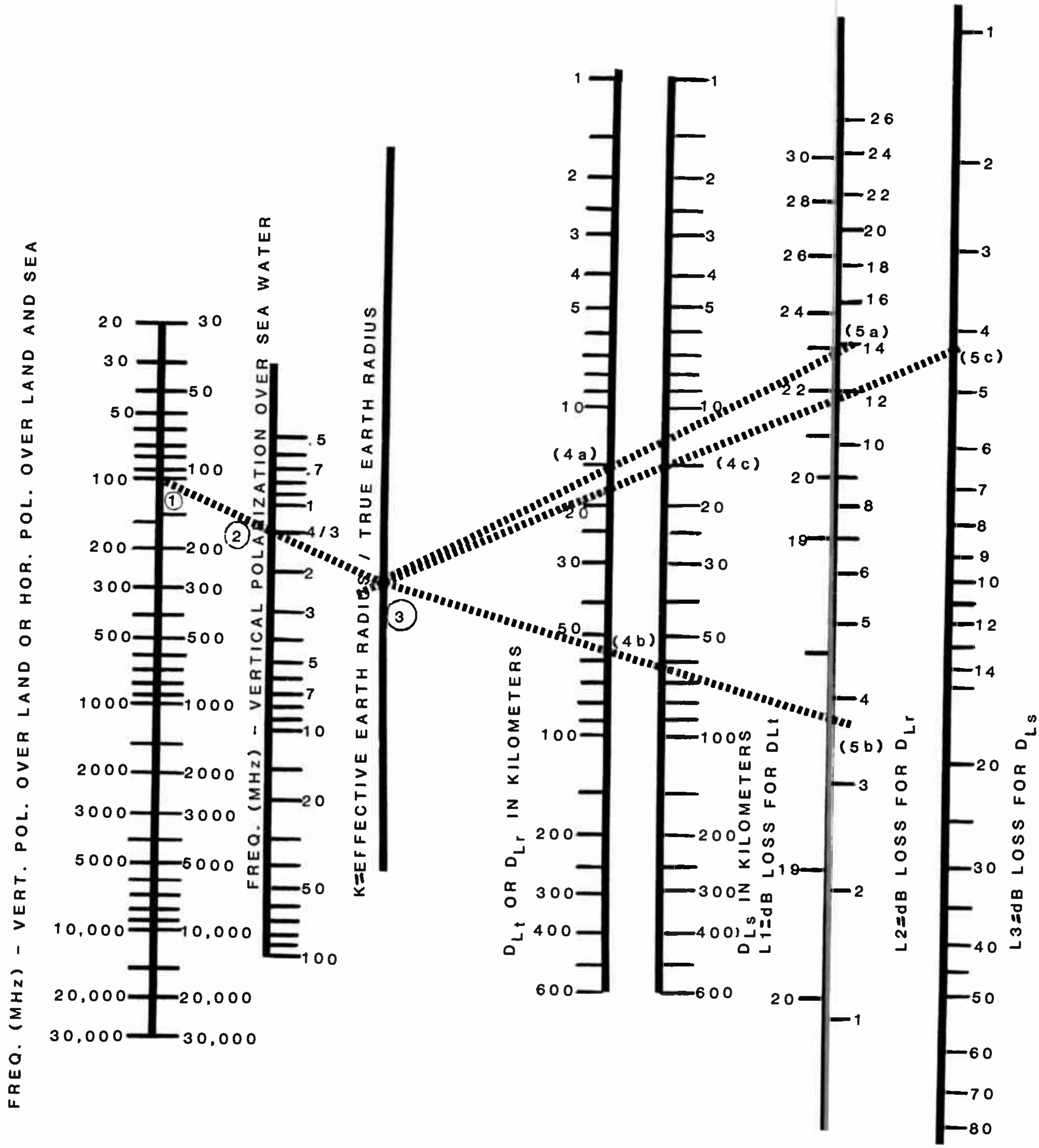


Figure 2.1-8. Diffraction loss over smooth earth.

h_{lt} = 14 m (14.9 ft); h_{lr} = 178 m (583.8 ft); F = 100 MHz; k = 4/3; total path length of 85 km; and the wave is vertically polarized over land. The distances are calculated to be: d_{lt} = 15 km (9.3 miles), d_{lr} = 55 km (34.2 miles), and d_{ls} = 15 km (9.3 miles). The total loss relative to free space is this L = L1 + L2 + L3 = 22.9 + 3.8 + 4.3 = 31.0 dB.[7]

Effects of Obstacles on Propagation

In the previous sections, a perfectly smooth sphere was assumed for earth. Only the effects of the atmosphere were accounted for in the k factor. The assumption of a perfectly smooth earth allowed for a relatively simple calculation of the expected field strengths and transmission losses at various points within the line of sight and regions beyond the line of sight. However, the real world is much less than ideal, and the presence of hills, buildings, foliage, as well as the atmosphere all have a bearing on the computation of field strengths. These obstacles have a complex effect on the propagation of radio waves which makes it virtually impossible to predict the field strength or transmission losses at discrete points close to these obstacles. However, the path being considered may be quantized by use of earth profiles, and through the use of some simplifying assumptions, predictions of the field strength which are more accurate than smooth earth approximations can be performed.

Hills

Perhaps the most common obstructions that will appear in the path of a radio wave are hills. The amount of attenuation the hill will introduce into a path is a function of the distance from the antenna terminals to the hill, and the height of the hill above or below the line-of-sight ray between the transmitting and receiving antennas. The hill's height and distance from the antenna can be determined by constructing a path profile and plotting the terrain features on special graph paper that includes the effect of refraction. The most common charts are defined for a factor of k of 4/3. A typical path is shown in Figure 2.1-9. Terrain elevations necessary to construct a path profile can be obtained from topographical maps of the area, or from computer databases. To obtain the best accuracy in using topographical maps the smallest scale available should be used, 1:24,000. These maps can be obtained from the United Stated Geological Survey (USGS) in Denver, Colorado or from any of its branch offices. Computer models exist that can map the terrain along a propagation path and are available from many sources.

In order to determine when a hill is sufficiently removed from a path to allow free-space conditions to exist, the Fresnel zone clearance equation can be used. This equation was initially developed to explain the diffraction of light around knife edged obstacles, and has since been applied to radio theory. This equation describes a radio path as an ellipsoid with the transmitting and receiving antenna located at the focal points of the ellipse. As Figure 2.1-10 depicts, the curves for various reflection coefficients intersect at 0 dB from free space when the clearance is equal to six-tenths of the distance to the first Fresnel zone clearance. Thus free space conditions exist when obstacles are outside the 0.6 F_1 zone radius. This distance can be calculated by:

$$h = 0.6\,F_1 = 328.6\sqrt{d_1 d_2 / Fd} \qquad [21]$$

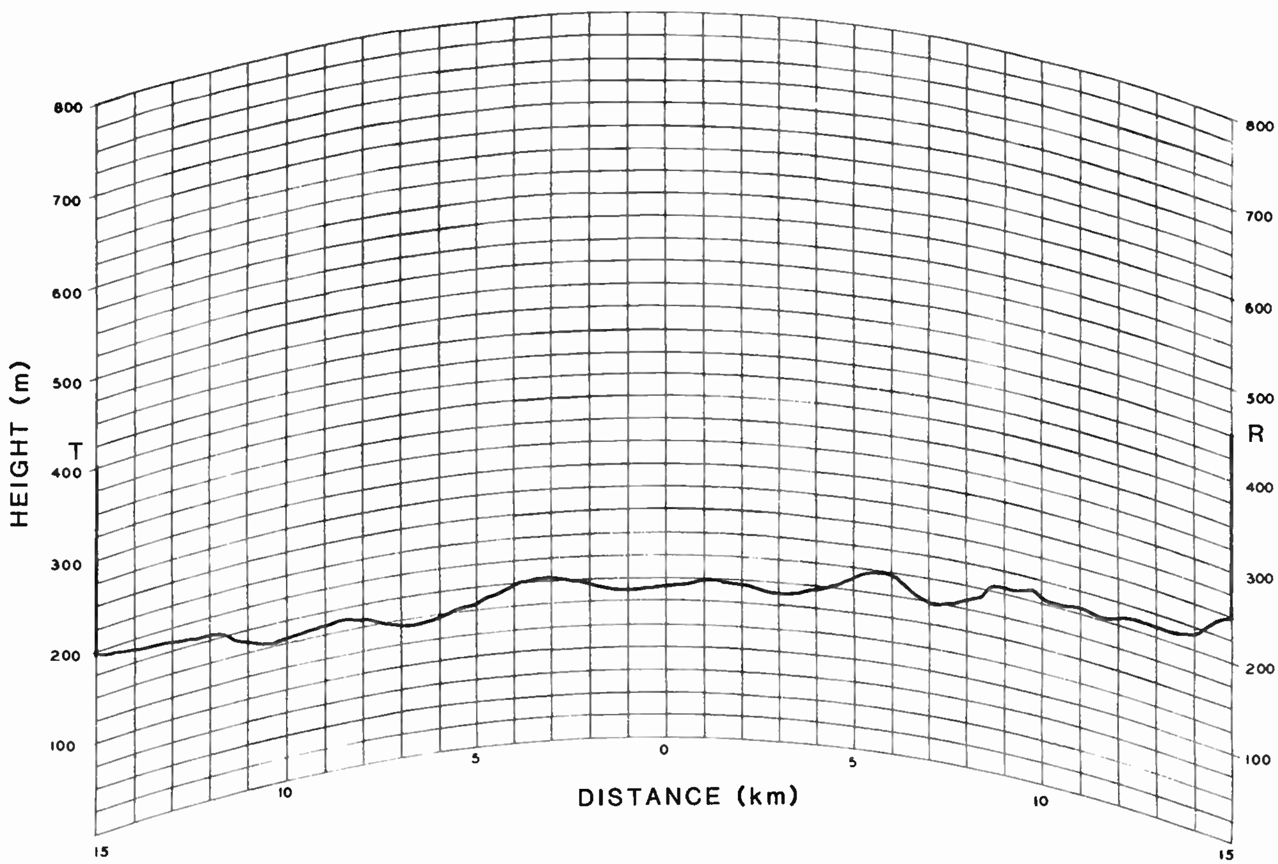


Figure 2.1-9. Typical 4/3 earth path profile.

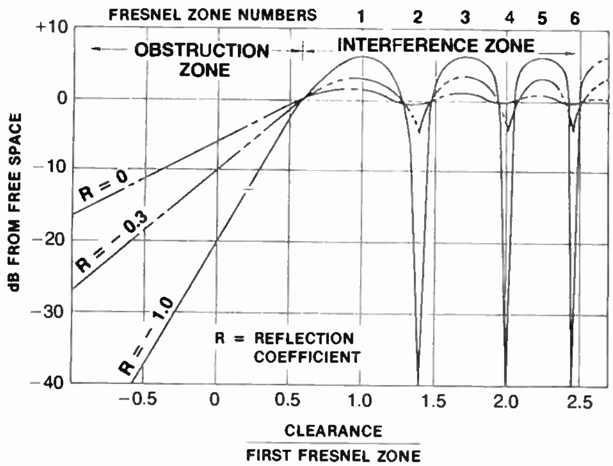


Figure 2.1-10. Effect of path clearance on radio propagation.

where the height of the 0.6 F_1 zone is in meters; d is the distance from one antenna to the obstacle in kilometers; d_2 is the distance from the second antenna to the obstacle in kilometers; d is the total path distance in kilometers; and F is the frequency in megahertz. When determining whether a path clears a hill, an additional height typically 15 meters, should be added to the height of the hill to account for any trees that may be present.

If the hill lies within the calculated Fresnel zone radium then nonfree-space conditions exist and additional losses will be present. When the frequency is high enough for the hill to be considered as a sharp ridge, and the transmitter and receiver are far enough away from the hill, then the loss may be calculated using diffraction from a knife edge, as shown in Figure 2.1-11. The height of the hill H(m) is measured from the line joining the centers of the two antenna to the top of the ridge. The amount of attenuation or shadow loss with respect to free space may be read from the graph shown in Figure 2.1-12. The value of v, the diffraction parameter, can be calculated, with respect to the distances measured in kilometers and the frequency F(MHz) from:[4]

$$v = 0.00258H\sqrt{dF/d_1d_2} \qquad [22]$$

When considering paths that are obstructed by hills that appear rounded rather than knife edged, the attenuation can be calculated using diffraction around a cylindrical surface, as depicted in Figure 2.1-13. This condition

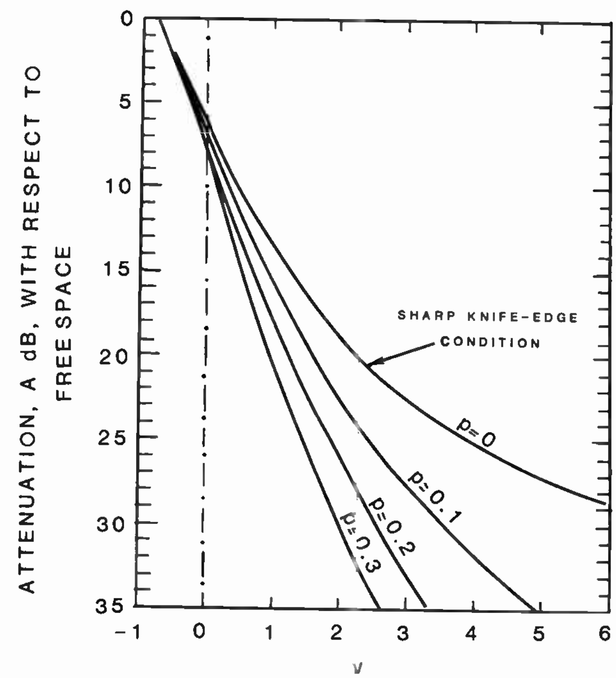


Figure 2.1-12. Attenuation due to various diffraction conditions.

predominates when the elevation of the hill changes drastically within a wavelength. This can occur when considering paths at the lower end of the VHF spectrum that pass over older mountain ranges, (Appalachian, Blue Ridge, Catskill).

The amount of attenuation can be found from the chart of Figure 2.1-12. The term p from the chart is a dimensionless quantity known as the index of curvature of the cylinder's radius $\boldsymbol{R}$ and is calculated from:[4]

$$\rho = 0.83\boldsymbol{R}^{1/3}\lambda^{1/4}\sqrt{d/d_1d_2} \qquad [23]$$

where all distances are in the same units. For those interested in incorporating the calculation of losses due to diffraction over knife edge and rounded obstacles into computers more exact equations can be found in Rice, et al.[9]

While the method for calculating the loss due to a single obstacle is relatively straightforward, there are times when successive obstacles are present in a radio path as shown in Figure 2.1-14. In order to determine the loss associated with multiple diffraction regions, an approximation method has been developed based on

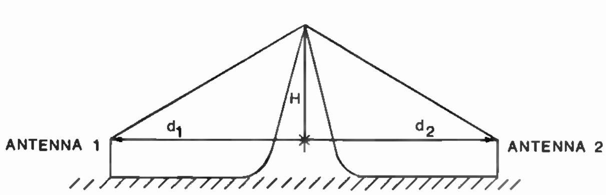


Figure 2.1-11. Ray path for knife edge diffraction.

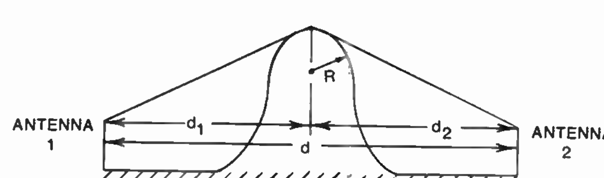


Figure 2.1-13. Diffraction due to cylinder.

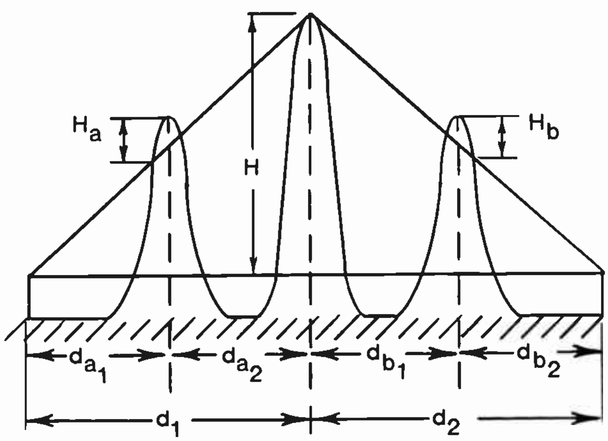


Figure 2.1-14. Multiple knife edge diffraction.

an extension of single-edge diffraction. The obstacles which would singly produce the greatest diffraction loss is determined using the methods discussed above. Lines are drawn joining the summit of this obstacle to the transmitter and receiver antenna locations. The additional attenuation caused by the remaining obstacles should be calculated using their heights, H_a and H_b, above these lines. These additional losses are then added to the loss due to the main obstacle in calculating the nonfree-space loss. It is important to note that even if H_a and H_b are slightly negative (below the lines) they may still produce a small amount of attenuation due to the Fresnel zone clearance requirements.[4]

Buildings

When planning for transmitting locations within built-up areas of cities or residential areas, buildings will have an effect on radio propagation. For radio relay stations, such as studio-to-transmitter links it is the normal practice to select sites that will be clear of buildings. However, where this is not feasible and the path geometry is known, (height and location of buildings) then the diffraction methods discussed for hills may be applied. In planning for broadcast systems it is not practical to relate attenuation measurements made in built-up areas to the particular geometry of buildings. Therefore it is more conventional to treat the losses in statistical manner, dividing the general classifications of building types into loss groups, so that a loss can be derived for a particular type of building (multi-story made of concrete and steel versus single story residential made of wood).

Within built-up areas there is much more back scatter than in open country. Additionally, due to the fact that buildings are more transparent to radio waves than the earth, there tends to be less shadow loss caused by buildings. However, the angles of diffraction due to buildings are usually much greater than in open country for natural terrain and thus the loss resulting from the presence of buildings tends to increase. Measurements indicate that at 100 MHz the median field strengths are 4 dB to 6 dB below that expected for a plane earth and drop off to about 10 dB for 200 MHz.[3] These measurements were made in areas containing some large buildings and open areas, but mainly consisting of residential areas. Measurements conducted in the 850 MHz band indicate field strengths 20 dB to 34 dB below that expected for free space for path distances of 1 km to 25 km.[10]

Vegetation

Among the many factors that have an effect on the determination of the losses present in a propagation path, vegetation is sometimes the most overlooked. Depending on the type of terrain in consideration, (open or forest) the effect of vegetation can add a several dB loss to the system. The amount of attenuation present is dependent upon the frequency and polarization of the wave (see Figure 2.1-15). As can be seen, the attenuation for a horizontally polarized wave for frequencies below about 1000 MHz is much less than for a vertically polarized wave. At around 1000 MHz, trees that are thick enough to block the field of vision can be modeled as an almost solid obstruction and the attenuation over or around these obstructions can be predicted from knife edge diffraction methods.[3]

The effect of vegetation on a radio path varies seasonally in the case of deciduous trees. During the winter months the losses due to shadowing and absorption are less than those during the spring and summer. It is interesting to note that the greatest losses will occur during the spring since new growth has more sap and moisture content which adds to the absorption losses. When the antenna is raised above trees and other vegetation, the prediction of field strengths depends upon the estimation of the height of the antenna above areas of reflection and the reflection coefficients.

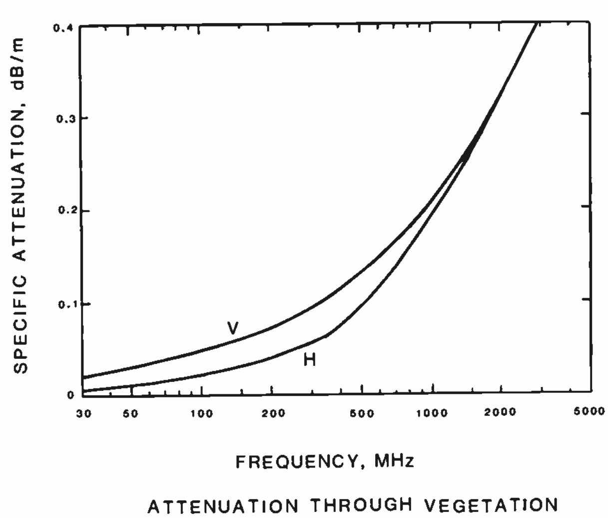


Figure 2.1-15. Attenuation through vegetation.

For areas of fairly uniform growth and for angles of incidence approaching grazing, the reflection coefficient will approach −1 at about 30 MHz. Even low growth that is uniform, (a wheat field) may yield a value of −0.3 for the reflection coefficient.

Atmosphere

As was discussed earlier, the troposphere is the major medium for propagation at VHF frequencies. The refractive index (n) of air has a value near unity (typically 1.00035). The index is dependent upon the dielectric constant, and can vary depending on the pressure and temperature of the air and on the amount of water vapor present. Therefore the refractive index changes with weather conditions and with the height above the earth. The velocity of radio waves is dependent on the refractive index of the atmosphere. As a general rule the velocity of a wave is slower at the earth's surface than at higher altitudes. So a horizontally polarized wave will be refracted back towards the earth, though unusual atmospheric conditions may change this, some simplifying assumptions are generally needed to obtain a solution under known meteorological conditions.

Ducting

Changes in the index of refraction of only a few parts per million can have dramatic effects on radio waves. Therefore it is usually more convenient to refer to the refractive index in terms of the refractivity, N:

$$N = (n - 1) \times 10^6 \qquad [24]$$

Under meteorological conditions where the refractive index decreases rapidly with height over a large horizontal distance, radio waves can become trapped and experience low propagation loss over long distances. This phenomenon is known as ducting. Although ducting is frequent with some locations and meteorological conditions, due to its randomness and long range unpredictability it is not a reliable mode for communications. However, due to the strong fields over the horizon caused by ducting, inter-station interference can result. In addition, line of sight paths may be affected by severe fading.

In order for atmospheric ducts to occur, two conditions must exist. First the refractive index gradient must be equal to or more negative than −157 N/km. The refractive index gradient is a measure of the change of the refractivity across a vertical height h, dN/dh. When this condition is present, the radio waves will remain close to the earth's surface beyond the normal horizon. Secondly, the refractive index gradient must be maintained over a height of many wavelengths. The duct may be thought of in the same manner as a transmission line waveguide. However unlike metallic waveguides, natural ducts do not have sharp boundaries, although there is a wavelength cut-off above which waves will not propagate. Since the duct does not have sharp boundaries, the thickness (t) will not be rigid. Therefore the cut-off wavelength (λ) will not be fixed but an estimate can be obtained from:[4]

$$\lambda = 2.5 \times 10^{-3} t^{2/3} \sqrt{\frac{\delta N}{t} - 0.157} \qquad [25]$$

where the wavelength and thickness are in meters. The term δN represents change in refractive index across the duct. As an example, a duct near the ground that is 25 meters thick and has a refractive index change of 10 N, (400 N/km) will have a cut-off wavelength of 0.15 m (2 GHz). However, a duct with the same refractive index gradient will have to be about 87 meters thick to propagate a wavelength of 1 meter (300 MHz).

A duct spreads the energy within it in the horizontal direction, but is contained in the vertical direction as the distance from the transmitter is increased. Thus, in principle it is possible for the field strength within a duct to be greater than the free-space field at the same distance. However, a duct will leak, or allow energy to escape at the boundary, adding to the transmission losses so that field strengths are seldom greater than free-space values.[4]

There are typically two types of ducts: ground based and elevated. A ground-based duct forms close to the earth's surface. Energy is propagated in this duct by being refracted back to the earth, reflected off the earth, then refracted again, see Figure 2.1-16(a). An elevated duct forms above the earth's surface and is generally very short lived. Energy in an elevated duct is refracted back and forth between boundaries without coming in contact with the earth, similar to the way coherent light propagates in a graded index optical fiber. (See Figure 2.1-16(b).) Shadow regions are formed along the area outside of a duct where, due to the nature of the duct, radio waves are not present. Receiving antennas placed in such a region will experience a loss

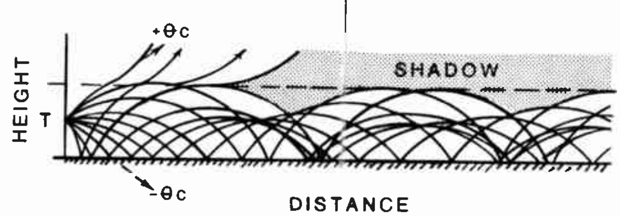


Figure 2.1-16a. Ray propagation in a ground based duct.

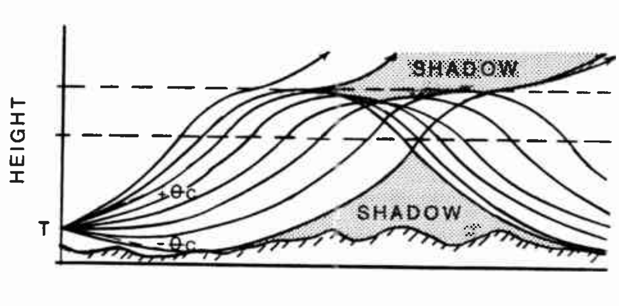


Figure 2.1-16b. Ray propagation in an elevated duct.

of signal. As can be seen from Figures 2.1-16(a) and 2.1-16(b) these regions can form not only above the earth's surface from a ground-based duct, but can also form along the earth's surface in the case of an elevated duct. Therefore a shadow region, that can result in loss of communications, can form at a receiver that is located relatively close to the transmitter.

Radio waves that leave the transmitting antenna at an angle greater than a certain angle, the critical angle, will not become trapped in a duct. These radio waves will propagate through the boundary of the duct, though they will experience some bending due to the change in the index of refraction at the duct's boundary.

Atmospheric Absorption

Radio systems using frequencies above 1 GHz experience another loss that must be accounted for when planning the system: atmospheric absorption. Relay links and STLs that are inadequately engineered may experience outages during periods of heavy rainfall due to this loss. The amount of attenuation due to rain is dependent upon three factors: (1) the rate the rain is falling; (2) frequency of the wave and (3) length of the rain cell a wave must propagate through. If the path length is only several kilometers long, it is usually adequate to approximate the length of the rain cell by the total path length. The average rainfall rate varies from one section of the country to another, however typical rainfall rates are given in Table 2.1-3. The specific attenuation

$$\gamma_r = KR_r^\alpha dB/Km \qquad [26]$$

where R_r is the rainfall rate in millimeters per hours, and the terms K and α are found[9] as:

$$K = [3(F-2)^2 - 2(F-2)] \times 10^{-4} \qquad [27.1]$$

$$\alpha = [1.14 - 0.07(F-2)^{1/3}] \times [1 + 0.085(F-3.5)e^{(-0.0066F^2)}] \qquad [27.2]$$

where F is the frequency in GHz. These equations give a good approximation to attenuation curves published by the ITU-R for frequencies below 50 GHz.[14]

In addition to rain, attenuation can also be caused by water vapor and oxygen that is present in the air. The attenuation due to water vapor and oxygen is less than that for rain and usually can be neglected. However, as radio systems employ higher microwave frequencies, attenuation due to these losses will become significant.

Table 2.1-3
Rainfall Amounts

Characteristics	Rate
Drizzle	0.25 mm/hr
Light Rain	1.00 mm/hr
Moderate Rain	4.00 mm/hr
Heavy Rain	16.00 mm/hr
Very Heavy Rain	100.00 mm/hr

Coverage Areas

Engineering a radio or television broadcast station using the methods presented previously are too cumbersome to be of any practical use in determining the service area of the station. While radio waves actually behave in the manner described in the previous sections, it would be too involved to use these methods at every point surrounding a station. Therefore other quantitative methods are needed to determine field strengths quickly and reliably. Considerable work has been conducted in this area and is still being carried out.

As can be seen from previous sections, the received field strengths are subject to natural and manmade phenomena. These can cause the field strengths to vary over periods of time and from one location to another. These changes can be long term such as seasonal change (weather, temperature and foliage) or short-term changes such as weather disturbances (storms and fronts, and vehicles passing in front of the receiver). These variations have an effect on radio systems that is difficult to account for when determining service areas. Thus it is appropriate to describe the field strength statistically, by what percentage of locations will receive a particular field strength for what percentage of time. By describing field variations in this manner, it is possible to determine the service area of a station. However, the terrain still needs to be defined. In preparing propagation curves, this is accomplished by incorporating a terrain roughness factor h. The terrain roughness factor is a generalization of the local terrain and is defined as the difference in elevation between the levels exceeded for 10% and 90% of the terrain along a path. The average value of h for the United States is 50 meters.[11] In using the propagation curves found in the FCC Rules and Regulations for FM and television stations, the local terrain is accounted for by determining the height of the antenna above average terrain along a radial.[12]

Field Strength Prediction

To simplify field strength prediction, curves have been developed to determine the service area of a station. These curves are generally developed using measured values taken from different geographical areas over certain periods of time. The median values are incorporated into a family of curves that describe the field strengths for various antenna heights, frequencies, and distances. The curves used by the FCC for FM and NTSC TV describe the field strengths for service at 50% of the locations for 50% of the time. These curves are referred to as F(50,50) and are based on an effective power of 1 kW radiated from a half-wave dipole in free space. The F(50,10) curves used by the FCC describe the field strength for 50% of the locations for 10% of the time. These curves can be used in conjunction with the method described by Allen to estimate the service provided by FM and television stations.[3]

Computer Databases

Through the use of personal computers, field strength estimates can be made quickly, allowing designers to try more options and see the effect on the service area. The designer can change transmitter locations, power levels and tower heights to optimize the station. There are typically four modules which comprise a computer simulation package: 1) the environmental data base, 2) the equipment data base, 3) the propagation loss module, and 4) the graphical output module.

The environmental data base defines the conditions in which the radio station must operate. For stations operating in the AM broadcast band this database is a digitized version of ground conductivity constants for an area. Programs are available which allow a user to search an area to determine appropriate ground conductivity values. For stations operating in the frequency range above 30 MHz this is a topographical database. The USGS and National Geophysical Data Center (NGDC) provide digitized terrain elevation data. The most common resolutions provided for terrain elevation provides data spaced at 3 or 30 arc-second intervals. The 30 arc-second database still provides acceptable results for area of coverage predictions but should be used with caution for STL or microwave planning as it may underestimate significant terrain features. For detailed point-to-point profiling, the 3 arc-second database is recommended.

The equipment database defines variables relating to equipment used in the radio system. This includes transmitter output, receiver sensitivity (or field strength), and feed line losses. More sophisticated packages will allow the user to define additional losses such as foliage and buildings to be accounted by the propagation loss program. In addition, these packages may allow a user to define additional standard deviation values to Rayleigh fading losses. Many packages account for transmitter antennas in varying degrees of sophistication. The simplest accounts for only omnidirectional patterns. More in-depth packages allow the user to define directional antenna patterns.

The propagation-loss module uses the information contained in the environmental and equipment databases to calculate the propagation loss for the area. There are two typical methods for calculating the propagation losses, theoretical and empirical. Theoretical methods are based on well understood physical equations as presented earlier in the chapter. These methods typically do not account for such clutter factors as foliage and buildings. These losses must be accounted for by the user. Empirical methods are based on measured observances of actual propagation characteristics for a particular frequency band and geography. These methods typically employ measurements taken throughout a range of environments, and hence account for such factors as foliage and buildings.

The graphical output module presents the data derived from the propagation loss module in a form easily understood by the user.

REFERENCES

1. *IEEE Standard Dictionary of Electrical and Electronic Terms,* ANSI/IEEE Std 100-1977, Wiley-Interscience.
2. *Reference Data For Radio Engineers,* 6th ed., 1982, Howard W. Sams & Co.
3. Allen, Edward W., Wave Propagation, Radiation, and Absorption, *NAB Engineering Handbook,* 6th ed., p. 119, 1975.
4. Hall, Martin: *Effects of the Troposphere on Radio Communications,* 1979.
5. Jordan, Edward C., *Electromagnetic Waves and Radiating Systems,* Prentice-Hall, Inc., Englewood Cliffs, pp. 608–688, 1950.
6. Norton, K.A., "Ground Wave Intensity Over a Finitely Conducting Spherical Earth," *Proceedings of the IRE,* p. 623, December, 1941.
7. Bullington, K. "Radio Propagation Variations at VHF and UHF," *Proceedings of the IRE,* p. 27, January, 1950.
8. "Wave Propagation," *The ARRL Antenna Book,* 14th ed., American Radio Relay League, pp. 1–19, 1982.
9. Rice, Longley, Norton, and Barsis, "Transmission Loss Predictions for Tropospheric Communication Circuits," *National Bureau of Standards Technical Note 101 (Rev.)*
10. Okumura, "Field Strength and its Variability in VHF and UHF Land-Mobile Radio Service," Rev. Elec. Com. Lab. 16, Tokyo, pp. 825–873, 1968.
11. Damelin, Daniel, Fine, and Waldo "Development of VHF and UHF Propagation Curves for TV and FM Broadcasting," FCC Report No. R-6602.
12. Federal Communications Commission, Rules and Regulations, Section 73.
13. Wang, John C.H., "A Skywave Propagation Study in Preparation for the 1605–1705 kHz Broadcasting Conference," *IEEE Transactions on Broadcasting,* vol. BC-31, pp. 10–17, March 1985.
14. "Attenuation and Scattering by Rain and Other Atmospheric Particles," CCIR Report 721.

2.2
DESIGN, ERECTION, AND MAINTENANCE OF ANTENNA STRUCTURES

JOHN WINDLE, P.E.
CONSULTING STRUCTURAL ENGINEER, WEST CHESTER, PA

THOMAS J. HOENNINGER, P.E.
STAINLESS, INC., NORTH WALES, PA

TOWER LIGHTING SECTION BY:
LEWIS WETZEL
HUGHEY & PHILLIPS, INC., WENHAM, MA

INTRODUCTION

The purpose of this chapter is to provide broadcast engineers and managers information concerning the design, erection, and maintenance of antenna structures. While fundamental principles of the design and behavior of these structures will be discussed, this chapter is not intended to enable readers to design and build their own tower, but provides instead a basic understanding of these unique structures to facilitate planning, modifying, and maintaining broadcast facilities.

TOWER CHARACTERISTICS

All towers may be classified in one of the two basic groups, guyed or self-supporting. As their names imply, guyed towers depend on cables extending from the tower to anchors located some distances from the tower base for their structural integrity, while self-supporting towers rely solely on their own construction as a cantilevered space truss.

With only a few exceptions, the cost of the actual tower structure and foundations is considerably less for a guyed tower than for one that is self-supporting. The advantage of the self-supporting tower is relatively small land area required. Therefore, the choice between guyed or self-supporting depends to a large degree on the availability and cost of real estate.

A self-supporting tower requires a nearly square plot of land with equal sides that are 8% to 20% of the tower's height, provided local zoning rules do not require tower height radius of land.

The amount of land required for a guyed tower depends on the distance between the tower base and the guy anchors. This distance is preferably between 70% and 80% of the height, which would require a rectangular plot having sides equal to 125% and 145% of the height.

Because of the great flexibility in guyed tower design, it is possible to reduce the anchor distance to as little as 35% of height, thereby requiring a much smaller land area. However, the cost of the tower increases as the anchor distance decreases. The approximate relationship of cost to anchor distance for a representative 1,200 ft television broadcast tower is shown in Figure 2.2-1.

It is often possible to position a guyed tower on an irregularly shaped plot or to obtain long term lease agreements or easements for guy paths and anchor locations in order to minimize the tower cost without obtaining large, rectangular land areas.

Self-Supporting Towers

Self-supporting towers may be either square or triangular in cross section. While it is usually more economical to use a triangular cross section, there are situations

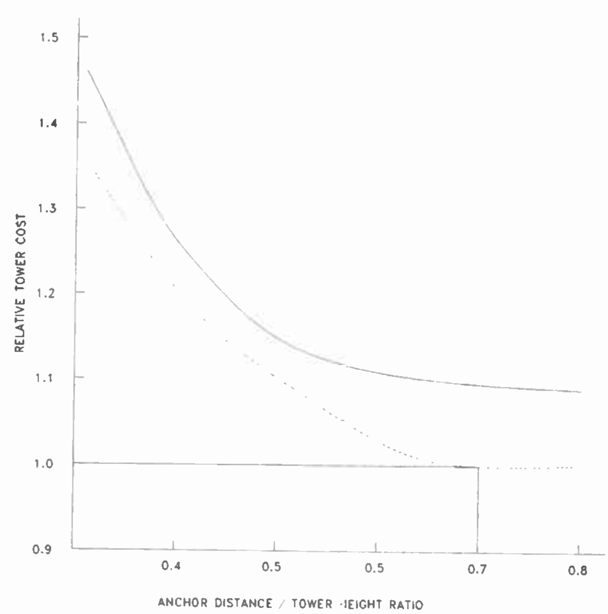


Figure 2.2-1. Effects of anchor distance on cost of a 1,200 foot guyed TV broadcast tower using ANSI/TIA/EIA Standard 222-F-1996.

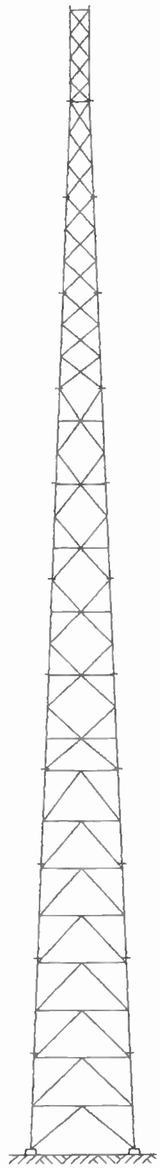


Figure 2.2-2. Elevation view typical self supporting tower.

where a square cross section is a better choice. The principle structural elements are the legs, the web bracing in each face, and if required for stability, horizontal diaphragm bracing. The legs are usually sloped (tapered) to provide adequate strength and stability as the height increases. The degree of slope is an option of the designer to suit the equipment supported, the required rigidity, and the available land area. The slope is sometimes varied within a tower to maintain a desirable balance between the costs of leg members and bracing, or to reduce the foundation loads. Frequently, the legs in the top section of the tower will be parallel to simplify the mounting of equipment (see Figure 2.2-2).

There are several different configurations of bracing members for the individual truss panels. The choice is influenced by the width of the panel, the magnitude of the wind and ice loads imposed, the location of equipment and required stability. Continuity in transferring the applied loads through the structure without significant eccentricity is essential regardless of the configuration used.

Guyed Towers

Guyed towers are almost always of triangular cross section although there are a few unique conditions for microwave and panel type FM and TV antenna supports where a square cross section is advantageous. The principle structural elements are the legs, the web bracing in each face, and the guy support systems (see Figure 2.2-3). Except for sections at the tower base and locations where the width changes, the legs are parallel. The width of the tower is usually constant throughout the height of the tower with the exception of sections supporting antennas requiring a specific width of support structure. The base section is often tapered to a single point to provide a pivot support to eliminate large bending and torsional moments.

Theoretically, there are an infinite number of arrangements of guy cables to support a tower. The most common arrangement is three cables spaced at 120° with one attached to each leg, as shown in Figure 2.2-4(a). This is the minimum number of cables that can be used. When the tower supports equipment, which imposes large twisting moments (torque), it is necessary to provide six cables at a level to maintain torsional stability. If the torque is localized, the guys at this location may be attached to triangular frames as shown in Figure 2.2-4(b). If the torque occurs throughout the height, it may be desirable to double-guy the tower at every level as shown in Figure 2.2-4(c).

The number of guy cable levels required to support the tower is dependent on a number of factors including the height of the tower, width, location of equipment and the environmental loading conditions. Because the tower is an axially compressed column, its strength is a function of its slenderness. While design codes permit

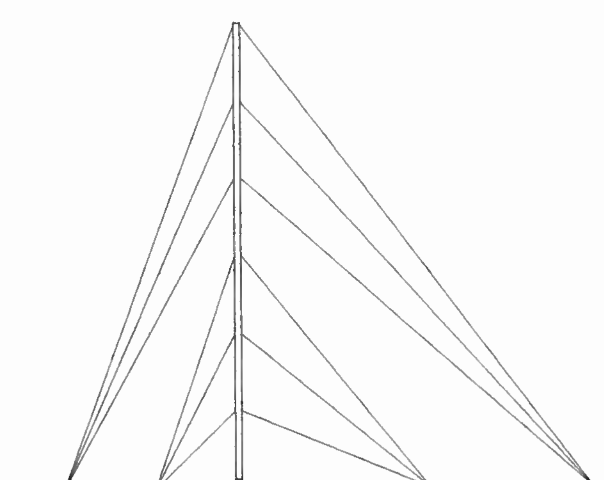


Figure 2.2-3. Elevation view typical guyed tower.

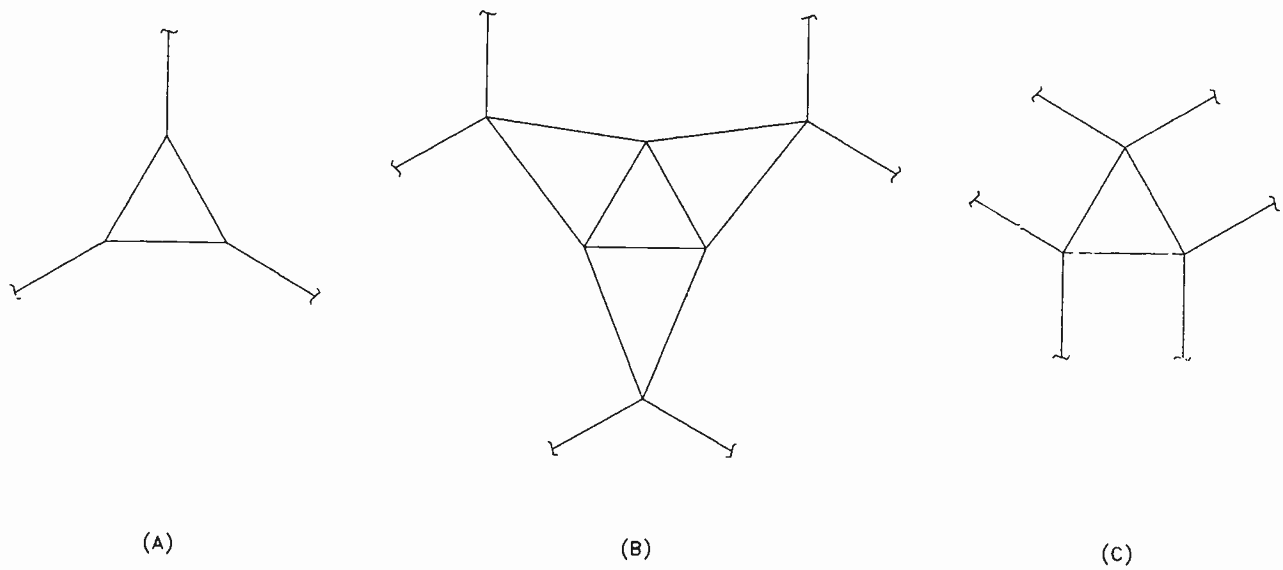


Figure 2.2-4. Typical guy arrangements.

slenderness ratios resulting in triangular towers having a span-to-width ratio as great as 49, it is usually economical to limit the ratio to a maximum of 30. While there is no upper limit to the number of guy levels imposed by any code, a practical limit for economical design is ten.

The position of equipment on the tower is an important factor in determining the location of guy levels. Preferably, guy attachments should not be located within the apertures of side mounted TV and FM broadcast antennas. Equipment producing large localized wind loads, such as microwave antennas or clusters of two-way radio cabinets and antennas, should not be positioned near the center of a span between guys.

If the tower will be subjected to ice loading, it is desirable to reduce the number of guy levels to minimize loads imposed on the tower by ice accumulation on the guy cables.

The number of anchors in each guy direction is dependent on several factors including the number of guy levels, the soil conditions, topography, and obstacles. As a general guideline, it is desirable to limit the number of guy levels attached to a single anchor to five. However, there is nothing absolute about this number, and other conditions may dictate using an anchor for a greater number. There are some soil conditions where it may be economical to provide two or more smaller anchors, while another instance the use of one large anchor might be desirable. If minimizing the area within which the tower would fall in the event of collapse is a consideration, a minimum of two anchors should be used in each direction. Where the elevations of the anchors differ from the tower base, it is desirable to vary the distance of the anchors from the tower base to maintain nearly equal initial tensions in the guy cables. Anchors higher than the tower base should be moved toward the tower; and anchors that are lower, away from the tower. The designer should specify the amount of movement.

Materials

Nearly all broadcast towers are made from steel because it provides the most economical structure. The selection of the grade and shape of steel is obviously an important design consideration.

Steel used for towers commonly has low carbon content with yield strengths in the range of 36,000 to 60,000 psi. These materials have good ductility and are suitable for welding. Some towers have been built using higher grade materials with yield strengths up to 100,000 psi, but the savings in weight are more than offset by higher base prices and increased fabrication costs. Regardless of the grade of material, the steel's mechanical and chemical properties should be certified by the producing mill to ensure that it conforms to the design requirements.

The shape of the material as well as its size and strength affects the tower's load carrying capacity. The shape also has a significant effect on the magnitude of loads produced by wind. Design standards permit a reduced wind load on round members as little as 57% of the wind load for flat or angular members of the same width. For this reason, solid round bars, round structural tubes and pipe are often used. This advantage in wind load is offset somewhat by increased fabrication costs, due to the necessity of welding plates to connect the various members.

There is no one grade or shape of materials that is best. The choice depends to a large degree on the preference of the designer and the type of fabricating facilities available.

A factor equally as important as the selection of the grade and shape of the structural steel is the design of the connections. For shop welded connections, the compatibility of the base and filler metals and required preheat temperatures must be considered. The procedures used must be qualified and the welders certified to use them. Inspection procedures should be compatible with the weld design.

Bolts for field assembly may be of various types. Usually those for the main load carrying members are high strength. If positive resistance to slippage of the connections is required, they should be designed as friction connections.

Guys

The most common material for tower guys is galvanized strand. This material has excellent strength and durability. Its structural elongation due to seating of the individual wires in the strand is small and can be almost entirely eliminated by pre-stressing the strand to 50% of its breaking strength at the factory. This should be performed for guys on tall towers with factory connected end fittings.

For guys on AM towers, and those close to FM and TV antenna apertures, a nonconductive material is sometimes desirable. Two such materials that have been used are Kevlar rope and fiberglass rods. When using these materials, careful attention must be given to protection against corona effects, fatigue, and deterioration due to exposure to ultraviolet light. Also, their elongation characteristics under load must be evaluated. They require delicate handling at all times.

Just as for the tower structure, the connections for the guys are as important as the guy material itself. Some of the most common connections are as follows:

- *Sockets* of forged or cast steel attached with molten zinc or epoxy resins develop the full strength of the guy. They are normally installed at the factory and proof loaded to 50% of the guy breaking strength. This type of fitting is most common for the larger guys used on tall towers.
- *Dead end grips* are preformed spiral wire loops in the shape of large hairpins. The two legs of the hairpin are wrapped around the guy with its closed end forming an eye. These grips are used for guys up to 1 in. in diameter and usually develop their full strength. They are easily installed in the field, but the ends must be completely snapped into place and a protective device installed to prevent ice from sliding down the guy and loosening the grip.
- *Clips* used to clamp the ends of guys (when properly applied and tightened) develop 90% of the guy's strength for sizes up to and including 7/8 in. and 80% for larger sizes. To install them it is necessary to bend the strand back on itself to form a loop; thus, the use of clips on large cables is difficult. The saddle of U-bolt type clips must be installed on the load side and not the dead end side, which provides another potential error in their installation.
- *Swaged sleeves* develop between 85% to 100% of a guy's strength depending on the size of the guy and equipment used to squeeze the sleeve. These fittings are usually installed at the factory and can be proof loaded. They are advantageous for connecting closely spaced insulators where the length of dead end grips is unacceptable.
- *Wedge Type Sockets* are available for guys up to 1-1/4 in. diameter and develop 100% of their strength. They are most advantageous for guys larger than those for which dead end grips are available.
- A *serving* is a connection made by rolling the individual wires of a strand back on the strand itself. This method has for the most part been replaced with dead end grips, but it is advantageous for small guys with closely spaced insulators.

Insulators

Insulators in radio frequency applications must withstand mechanical and electrical stress in a varied, changing exterior environment. Selection of insulators should be made with these factors in mind. Insulators primarily designed for 60 Hz applications are unsuitable, particularly at high RF powers.

The most common insulating material is a wet process porcelain, which has excellent compressive strength and good insulating capabilities for frequencies up to 2 MHz. Synthetic materials are also used. Some other types of insulators are as follows:

- *Base insulators* for AM towers are made from porcelain with appropriate steel end plates or ferrous/nonferrous castings. For guyed towers, a rocking arrangement is provided in the form of a convex plate and pin at the top, or a pivot pin at the bottom of the assembly to hold the tower in place and relieve the porcelain from bending loads which could cause cracking. For self-supporting towers, the insulators are bolted between the tower leg and base pier, and are designed to sustain both uplift and download while keeping the porcelain in compression.
- *Sectionalizing insulators* are sometimes required to isolate sections of a guyed tower. Where a compression load only is applied, a guyed tower base with minor modifications can be used. If a tension load is anticipated, a push-pull insulator similar to the type used for self-supporting towers is required. Under no circumstances should the porcelain be put in tension.
- *Guy insulators* are available for primary insulation and for break-up purposes. Primary insulation (insulators next to the tower) should be selected to withstand the full voltage appearing at the guy attachment point. This is to ensure that sufficient insulation remains if all the break-up insulators in the guy line flash over. Break-up insulators, used to reduce reradiation, are selected to withstand the transmitter induced voltage and static voltage. Break-up insulators are usually low voltage types, sometimes protected from flashover and subsequent power arc by a static dissipation device. Guy insulators are available in many styles classified as either compression or tension types.

Finishes

Steel is susceptible to deterioration from atmospheric corrosion. To prevent deterioration, the tower members and hardware must be given a protective coating. This coating is usually zinc, which has excellent resistance to corrosion, and, because it is higher in the electrochemical series of the periodic table of elements, it provides cathodic protection to exposed steel surfaces adjacent to it. Even though the zinc coating may be scraped or otherwise damaged, it continues to inhibit corrosion of these exposed areas, and rust will not develop beneath adjacent zinc coats.

There are several methods for applying the zinc including hot dip galvanizing, flame spraying, electroplating and painting. All must be applied to clean surfaces.

- *Hot dip galvanizing* consists of dipping the steel into a bath of molten zinc. A metallurgical bond develops between the steel and the zinc, which adheres to it. When galvanizing tubular members, it is necessary to provide holes in both ends to ensure that the inside surfaces are coated. Careful attention must be given to the type of base and weld metals used, as well as to the welding and forming procedures used in fabrication, to safeguard against possible embrittlement of the steel when galvanized. When properly applied, this process provides the most durable coating.
- *Flame spraying* consists of spraying molten zinc at high pressure onto the steel surfaces. The bond in this process is mechanical rather than metallurgical. The coating produced is more porous and has less resistance to abrasion than the hotdip galvanized coating. It cannot be used for the inside of hollow sections or other cavities where access is difficult.
- *Electroplating*, while suitable for small objects, does not produce a coating thick enough to withstand a hostile environment. This method is not recommended for tower parts or hardware.
- *Zinc rich paint* consists of extremely finely divided zinc in an inorganic or organic vehicle. It is not a metal coating method, but rather a painting procedure. Its resistance to abrasion and durability are less than hotdip galvanizing. This procedure is, however, useful for maintenance.

Ice Prevention

Coatings are available to reduce the adherence of water to surfaces and subsequently the formation of ice on them. However, no reliable means exists to completely remove the risk of severe ice accretion.

Access Facilities

A tower must have some access facilities in order to maintain it and the equipment the tower supports. For small towers, the bracing members of the tower itself often serve as steps, or step bolts are attached to one leg or face.

Ladders

For taller broadcast towers, a fixed ladder inside the tower is desirable. The Occupational Safety and Health Administration (OSHA) standards for these ladders require a minimum clear width between side rails of 16 in. and a maximum rung spacing of 12 in. OSHA also requires that any continuous ladder more than 20 ft in height be equipped with a safety device. This device consists of a continuous rail, either rigid or cable, running up the center of the ladder. A clamping device attached to the climber's safety belt rides along this rail. As long as the climber is in a normal position, the clamp slides freely; if the person begins to fall, a cam actuated mechanism freezes the clamp to the rail and prevents the person from falling.

Elevators

For tall towers supporting multiple antennas, it is often desirable to install an elevator. Most tower elevators are of the power, cable driven type with a capacity of 500 to 750 lbs. and a speed between 80 and 100 ft per minute. They consist of a drive mechanism, car, guide rails, hoist cable with supporting sheaves, tension weights, electronic controls and a two-way communications system.

Considerable attention must be given to elevator safety features. These should include limit switches to prevent travel beyond the upper and lower landings on the tower, an automatic brake on the driving mechanism that is activated by an interruption in power, a mechanism to automatically clamp the car to the rails in the event of a broken hoist cable and interlocks to prevent operation with the car gate open. It is advisable to determine the applicable state or municipal government regulations that may apply and whether permits, tests and inspections are required before the tower and elevator system are designed.

The added wind and dead loads from an elevator system are substantial and must be considered in the tower design. Also, careful attention must be given to the positioning of the ladder. RF transmission lines and electrical conduits in relationship to the elevator. The ladder must be positioned so it is accessible from the elevator car and can be used for an emergency descent. While the elevator hoist cables can be restrained in guides on the return side, they are free to move about under wind load on the lifting side. Therefore, the conduits and transmission lines must be protected from hoist cables striking and damaging them. If a side mounted TV or FM antenna produces a high RF field within the hoistway, protection must be provided to prevent arcing between the hoist cables, the tower structure, and other appurtenances.

Transmission Line Bridges

To allow for the different thermal expansion and contraction characteristics of broadcast towers and transmitter buildings, it is necessary to keep the tower separated from the transmitter building. Unless the transmission line is placed underground, it is necessary to provide a structural support for it at a height compatible with the transmitter location in the building. The top of the support can be covered with steel grating or plate to protect the line from falling ice. The details of this structure can become quite involved for sites with multiple antennas, uneven terrain and roadways, or obstacles between the tower and building.

Stairways

The lower landing for a tall, guyed broadcast tower with an elevator is often 30 ft or more above ground level. A stairway may be desirable to permit easier access to the landing. This structure can be combined with the transmission line support bridge, or it may be completely separate. It may also be desirable to install a small capacity boom above the lower landing to lift radio cabinets or other equipment onto the landing.

ANTENNA STRUCTURE REGISTRATION REQUIREMENTS

On November 30, 1995, the FCC adopted rules to streamline the Commission's antenna structure clearance process by requiring owners to register their antenna structures with the Commission by June 1998. The registration generally applies to structures more than 60.96 m (200 ft) above ground or located near an airport. Detailed information concerning registration is available from the FCC's Forms Distribution Center at 1-800-418-FORM or through the Internet at http://www.fcc.gov/wtb/antstruc.html.

Upon registering an antenna structure with the Commission, an owner will receive a Registration (FCC Form 854R) containing a Registration Number which uniquely identifies the structure. The owner must provide each tenant licensee and permittee with a copy of the Registration. In turn, licensees and permittees must reference this registration number on all subsequent Commission filings.

Marking and Lighting Antenna Structures to Meet FAA/FCC Requirements

Unless otherwise specified by the Commission, each new or altered antenna structure to be registered on or after July 1, 1996, must conform to the FAA's marking and lighting recommendations set forth on the structure's FAA determination of "No Hazard," as referenced in the following FAA Advisory Circulars: AC 70/7460-1H, *Obstruction Marking and Lighting* and AC 150/5345-43D, *Specification for Obstruction Lighting Equipment*. These documents contain FAA recommendations for marking and lighting structures which pose a potential hazard to air navigation. The Commission considers these specifications, standards and general requirements stated in these documents to be mandatory. The Advisory Circulars listed above may be obtained from the U.S. Department of Transportation, Subsequent Distribution Office, Ardmore East Business Center, 3341 Q 75th Avenue, Landover, Maryland 20785. Fax requests for these publications may be sent to (301) 386-5394.

CERTIFICATION PROCEDURE FOR FAA APPROVED LIGHTING EQUIPMENT

To assure that all lighting equipment manufactured to meet the specifications in the FAA Advisory Circulars, a certification program was established. On January 1, 1990, ETL Testing Laboratories, Inc., began administering the certification program for airport and obstruction lighting equipment. Lighting equipment approved by the FAA prior to January 1, 1990, continued its approved status without additional product qualification testing for a period of five years through a grandfather clause, provided sufficient documentation exists to support the certified status.

The FAA's Advisory Circular 150/5345-53, *Approved Airport Equipment* lists the certified lighting equipment that ETL has verified as meeting the applicable FAA specifications. When building or modifying an antenna structure that involves the use of obstruction lighting equipment, it would be wise to check the FAA approved lighting equipment list or require a certificate of compliance from the manufacturer.

FCC Lighting Requirements

During construction of an antenna structure for which obstruction lighting is required, a temporary light must be installed at the uppermost point of the structure. This light shall be similar in type and intensity to the permanent light that is required for that level. In addition, as the height of the structure exceeds each level at which permanent obstruction lights will be required, another level of temporary lighting must be installed. If practical, the permanent obstruction lights may be installed and operated at each required level as construction progresses.

Inspection of Antenna Structure Lights and Associated Control Equipment

The owner of any antenna structure which is registered with the Commission and has been assigned lighting specifications:

(a) (1) Shall make an observation of the antenna structure's lights at least once each 24 hours either visually or by observing an automatic properly maintained indicator designed to register any failure of such lights, to insure that all such lights are functioning properly as required: or alternatively. (2) Shall provide and properly maintain an automatic alarm system designed to detect any failure of such lights and to provide indication of such failure to the owner.

(b) Shall inspect at intervals not to exceed three months all automatic or mechanical control devices, indicators, and alarm systems associated with the antenna structure lighting to insure that such apparatus is functioning properly.

Notification of Extinguishment or Improper Functioning of Lights

The owner of each antenna structure which is registered with the Commission and has been assigned lighting specifications:

(a) Shall report immediately by telephone or fax to the nearest Flight Service Station or office of the

FAA any observed or otherwise known extinguishment or improper functioning of any top steady burning light or any flashing obstruction light, regardless of its position on the antenna structure, not corrected within 30 minutes. Such reports shall set forth the condition of the light or lights, the circumstances which caused the failure, the probable date for restoration of service, the FCC Antenna Structure Registration Number, the height of the structure (AGL and AMSL), and the name, title, address, and telephone number of the person making the report. Further notification by telephone or fax shall be given immediately upon resumption of normal operation of the light or lights.

(b) An extinguishment or improper functioning of a steady burning side intermediate light or lights, shall be corrected as soon as possible, but notification to the FAA of such extinguishment or improper functioning is not required.

Recording of Structure Light Inspections in the Owner Record

The owner of each antenna structure which is registered with the Commission and has been assigned lighting specifications must maintain a record of any observed or otherwise known extinguishment or improper functioning of a structure light and include the following information for each such event:

(a) The nature of such extinguishment or improper functioning.
(b) The date and time the extinguishment or improper operation was observed or otherwise noted.
(c) The date and time of FAA notification, if applicable.
(d) The date, time and nature of adjustments, repairs or replacements made.

Time When Lights Should be Exhibited

Red obstruction lights may be operated by a control device adjusted so the lights will be turned on when the northern sky illuminance reaching a vertical surface falls below a level of 60 footcandles (645.8 lux) but before reaching a level of 35 footcandles (376.7 lux). The control device should not turn the lights off until the northern sky illuminance rises to a level of at least 60 footcandles (645.8 lux). The lights may also remain on continuously.

Medium intensity white and dual obstruction lights should be operated by a control devise so the lights will switch from their day/twilight to night mode of operation when the northern sky illuminance reaching a vertical surface falls below a level of 5 footcandles (53.8 lux) but before reaching a level of 2 footcandles lux). The control device should reverse intensity changes when going from night to day/twilight mode.

High intensity white and dual obstruction lights should be operated by a control device so the lights will switch from their day to twilight mode of operation when the northern sky illuminance reaching a vertical surface falls below a level of 60 footcandles (645.8 lux) but should occur before it drops below 35 footcandles (376.7 lux). These lights should switch from their twilight to night mode of operation when the level falls below 5 footcandles (53.8 lux) but should occur before reaching a level of 2 footcandles (21.5 lux). The control device should reverse intensity changes when going from night to twilight to day within these same levels.

Options for Marking and Lighting Antenna Structures

Red obstruction lights are used to increase conspicuity during nighttime. Painting is required for daytime and twilight marking. The red obstruction lighting system is composed of flashing omnidirectional beacons (L-864) and steady burning lights (L-810). When the system is comprised of more than one level of flashing beacons, the beacons should flash simultaneously.

Medium Intensity Flashing White Lighting System

Medium intensity flashing white (L-865) obstruction lights provide conspicuity both day and night and eliminates the need to paint the structure. The medium intensity flashing white light system is composed of flashing omnidirectional lights. These systems are used on structures up to 500 ft (153 m).

Dual Red and White Flashing Medium Intensity Lighting System

The dual lighting system includes red flashing beacon or beacons (L-864) and steady burning red lights (L-810) for nighttime and medium intensity flashing white lights (L-865) for daytime and twilight use.

These systems may be comprised of a red flashing beacon (L-864) and a medium intensity flashing beacon (L-865) or a single beacon which incorporates both the red and white flashing beacons (L-864/L-865). These systems are used on structures up to 500 ft (153 m) and eliminate the need to paint the structure.

High Intensity Flashing White Lighting System

Lighting with high intensity flashing white obstruction lights (L-856) provides the highest degree of conspicuity both day and night. When this system is operated 24 hours a day, other methods of marking and lighting may be omitted. These systems are used on structures from 500 ft. (153 m) to the maximum height allowed by the FAA.

Dual Red and White Flashing High Intensity Lighting System

The dual lighting system includes flashing red lights (L-864) and steady burning red lights (L-810) for nighttime and high intensity flashing white lights (L-856) for daytime and twilight use. When using this system, marking the structure with paint is eliminated. These systems are used on structures from 500 ft (153 m) to the maximum height allowed by the FAA.

Other Electrical Circuits to Consider in Antenna Structure Planning

During the planning and design stages for an antenna structure, there are other electrical circuits that are necessary or should be considered to make its operation more efficient and provide a source for future income. Most tall antenna structures require deicing circuits with their associated control devices. An ac utility circuit can provide access to 120 volt power at selected elevations on the structure and will reduce costly maintenance time. If the structure's height justifies an elevator, control circuits for the elevator are necessary.

Circuits to provide power and multi-pair cables to various platforms are needed for rental communication system customers.

DESIGN STANDARDS

The vast majority of towers in the United States have been designed in accordance with the EIA-222, *Structural Standards for Steel Antenna Towers and Antenna Supporting Structures.* This standard has been used since 1959 when it replaced the Radio Electronic Television Manufacturers Association (RETMA) Standard TR-116. The current revision "F" of EIA-222 was issued in 1996. The standard has been approved by the American National Standards Institute and carries the designation ANSI/TIA/EIA-222-F-1996.

This standard is intended to provide minimum criteria for specifying and designing steel antenna towers and antenna supporting structures. Unlike general specifications and building codes, it is applicable only to antenna tower and supporting structures. As such it contains criteria specific to these structures that are not readily available elsewhere. Therefore, it is always advisable to specify that your tower must conform to this standard.

Appendix A: Purchase Checklist of this standard is provided to alert the purchaser to the most common area where site-specific data may be required to supplement the minimum criteria of the standard.

Statutory

Most municipal and state governments have statutory codes regulating the design of structures. Many of these are patterned after or include one of several model codes. The most common of these are:

- Building Officials and Code Administrator International (BOCA) Basic *Building Code.*
- International Conference of Building Officials (ICBO) *Uniform Building Code.*
- Southern Building Code Congress (SBC) *Standard Building Code.*

These codes cover all types of structures and are directed primarily toward conventional types of buildings. As such, they do not contain all the criteria necessary to design broadcast towers. For example, none of them includes a recommended safety factor for guy cables.

The industry standard ANSI/TIA/EIA-222-F-1996 is compatible with these codes. In fact, its use for calculating and applying wind loads is required by the BOCA Basic Building Code and, as an approved American National Standard, is permitted by the ICBO Uniform Building Code.

Since it is necessary to comply with the applicable statutory requirements, it is important to determine what these requirements are and include them in the purchase specifications for the tower.

Loads, Analysis & Safety Factors

In addition to a tower's own dead weight and the dead weight of the appurtenances and equipment it supports, the tower must withstand the forces of nature, wind, ice, temperature changes, and earthquakes.

Wind Load

Wind produces a principal load on tower structures. For design purposes it is represented as a horizontal static force.

Wind load is specified in terms of a basic wind speed at 10 m (33 ft) above ground level. The ANSI/TIA/EIA-222-F-1996 standard provides a tabulation of recommended minimum values for this speed for each county in the United States. This standard also gives specific procedures and factors for calculating wind loads considers the following:

- Wind pressure is proportional to the square of wind speed.
- Wind speed and, consequently, wind pressure vary with respect to the height above ground.
- The effects of gusts of brief duration, which exceed the fastest mile basic wind, speed.
- The effects of the configuration, size, proportions, shape and orientation with respect to the wind direction of the structural components of the tower and its appurtenances.

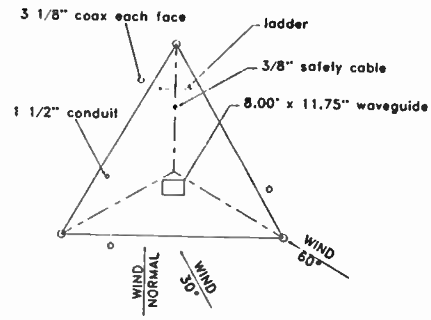
An example of wind load calculation for a typical broadcast tower is given in Figure 2.2-5.

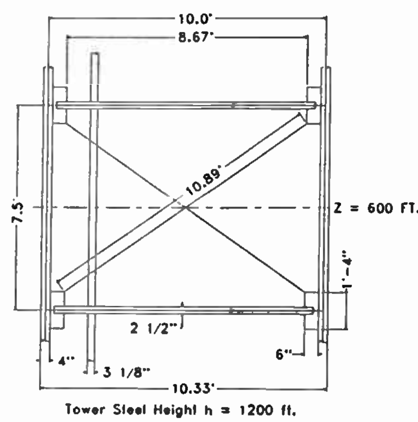
Since the wind may act from any direction it is necessary to apply the calculated wind loads in any horizontal direction to determine the maximum stresses produced in the structure. For a triangular tower, three directions must be considered; while for a square tower, two are sufficient (see Figure 2.2-6).

In addition to this direct load in the direction of the wind (drag), there may also be a component of load perpendicular to the wind direction (lift). These lift components are calculated in a manner similar to that for drag forces using different shape coefficients that vary with respect to the angle of attack between the member's geometric axis and the wind direction. They are most significant for wind acting on guy cables, microwave antennas and rectangular waveguides.

Ice Loads

Ice accumulations have two effects on a tower. The weight of the ice acts directly on the structure in the same manner as the dead weight. The ice accumulation





PROCEDURE

1. Determine basic wind speed V (mph). Minimum recommended values in Section 16
2. Calculate exposure coefficient $K_z = (Z/33)^{2/7}$ where Z is the height (ft) above ground level to midpoint of section. $1.0 \le K_z \le 2.58$.
3. Calculate velocity pressure. $Q_z = 0.00256 K_z V^2$ (lbs/sq.ft).
4. Calculate gust response factor. $G_h = .65 + 0.6/(h/33)^{1/7}$ $1.0 \le G_h \le 1.25$ where h is the height (ft) of the main tower structure.
5. Calculate the projected areas A_f and A_r (sq. ft.) of the flat and round structural components in one face including any linear appurtenances attached to a face and considered as structural appurtenances. Include the thickness of any ice load.
6. Calculate the solidity ratio $e = (A_f + A_r)/A_g$ where A_g is the gross area (sq. ft) of one tower face as if the face were solid.
7. Calculate the structure force coefficient. $C_f = 3.4e^2 - 4.7e + 3.4$ (Triangular cross section) $C_f = 4.0e^2 - 5.9e + 4.0$ (Square cross section)
8. Calculate reduction factor for round structural components $Rr = 0.51e^2 + 0.57 \le 1.0$.
9. Determine direction factors D_f and D_r from Table 2 for the wind direction considered.
10. Calculate the projected area A_e (sq. ft) of the structural components. $A_e = D_f \times A_f + D_r \times A_r$
11. Calculate the projected area A_a (sq. ft) of each individual appurtenance not considered as a structural component.
12. Determine the appropriate force coefficient C_a from Table 3 for each individual appurtenance and sum the products $C_a \times A_a$ (sq. ft).
13. Calculate total wind force. $F = q_z \times G_h (C_f \times A_e + \Sigma C_a \times A_a)$

EXAMPLE

1. V = 80 mph.
2. $K_z = (600/33)^{2/7} = 2.29$
3. $q_z = 0.00256(2.29)(80)^2 = 37.52$ lbs/sq. ft
4. $G_h = 0.65 + 0.60/(1200/33)^{1/7} = 1.009$
5.

	Structural Component	Af	Ar
Legs	2 x 4/12 x 7.5	-	5.00
Gusset PL	4 x (6 x 8)/144	1.33	-
Horizontal	2.5/12 x 8.67	1.81	-
Diagonals	2 x 0.75/12 x 10.84	-	1.36
3-1/8"coaxes	3.125/12 x 7.5	-	1.95
	Σ	3.14	8.31

6. $e = (3.14 + 8.31)/(10.33 \times 7.5) = 0.148$
7. $C_f = 3.4(0.148)^2 - 4.7(0.148) + 3.4 = 2.78$
8. $R_r = 0.51(0.148)^2 + 0.57 = 0.58$
9. For wind normal $D_f = 1.0$, $D_r = 1.0$.
10. $A_e = (1.0)(3.14) + (1.0)(0.58)(8.31) = 7.96$
11.

Appurtenance	A_a	C_a	$C_a \times A_a$
Ladder rails(2 x 0.75)/12 x 7.5	0.94	1.2	1.13
Lad. rungs(7.5 x 0.75)/12 x 1.33	0.63	1.2	0.76
Safety cable 0.375/12 x 7.5	0.23	1.2	0.28
1-½" Conduit 1.9/12 x 7.5	1.19	1.2	1.43
Waveguide 11.75/12 x 7.5	7.34	2.0	14.68

12. $\Sigma C_a \times A_a = 18.28$ sq.ft
13. $F = (37.52)(1.009)((2.78)(7.96) + 18.28) = 1531$ lbs.
$F = 1531/7.5 = 204.2$ lbs/ft.

Figure 2.2-5. Wind force calculations for latticed towers using ANSI/TIA/EIA-222-F-1996.

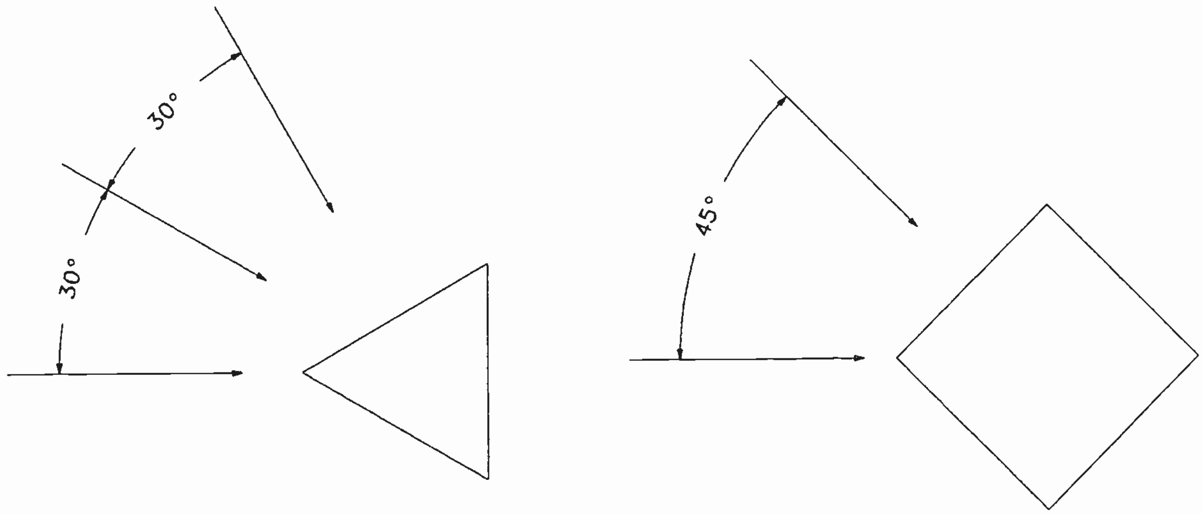


Figure 2.2-6. Wind directions to be considered.

also increases the area exposed to the wind and consequently the load produced by the wind. This increase is substantial on small components such as guy cables, tension rods, ladders, small diameter transmission lines and reflector screens for antennas. It is also possible for the ice accumulation to alter the aerodynamic shape of members, thereby requiring the use of a different coefficient in calculating the wind load. An example is a set of closely spaced parallel coaxial lines. Without ice each would be considered a round cylindrical member. With accumulated ice, they would present a large flat area to the wind requiring a different coefficient.

Ice produces an entirely different stress distribution in a tower than wind, so it is not reasonable to merely increase the design wind load to provide for ice accumulation. It is also a misconception that ice will break up and blow off the tower, and therefore, ice and wind need not be considered simultaneously. The ANSI/TIA/EIA-222-F-1996 Standard states that unless otherwise specified by the purchaser, 75% of the wind load shall be applied in combination with the ice load. This is equivalent to reducing the basic wind speed by a factor of 0.866. In areas where ice is likely to form, it is advisable to specify a basic wind speed to act concurrently with the ice load as well as the basic wind speed without ice), (70 mph with ice and 80 mph without ice).

While the ANSI/TIA/EIA-222-F-1996 Standard emphasizes the need to consider ice loads, it does not provide specific recommendations for the magnitude of the accumulation. This responsibility is left with the purchaser.

Temperature Changes

Changes in temperature have no significant load producing effects on self-supporting towers, but they can on guyed towers. Because of their differences in length, the guy cables expand and contract different amounts than the tower itself and thereby require elastic deformations from stress changes. The effects are greatest for those cables having the flattest angle with the ground. While the stresses produced are considerably less than those produced by wind and ice loads, they should be considered in the design of guyed tower.

Seismic Loads

Loads due to earthquakes are considered to act horizontally and are dependent on the mass and stiffness of the tower. They are usually less than those produced by wind but are distributed in a different manner. Procedures for calculating these loads are given in some of the design standards including ASCE 7, but not in ANSI/TIA/EIA-222-F-1996. While a tower properly designed for wind loads is usually adequate for seismic forces, they cannot be neglected in areas with frequent and intense earthquake occurrences.

Structural Models & Analysis

A self-supporting tower may be described structurally as a cantilevered space frame or truss. Although it may have many different members, it is a relatively simple structure, and the determination of the forces in the individual members due to the applied static loads is easily done using fundamental principles of structural mechanics. The potential modes of failure are buckling of individual leg or bracing members under compressive loads, and shear or tension failures of the connections.

A guyed tower is a much more complex structure than a self-supporting tower. Whereas there is only one basic path through a self-supporting tower for the loads to be transferred to the ground, there are several for a guyed tower. The distribution of the loads among these paths is dependent upon the relative stiffness of guy systems and the tower shaft.

Each span of the tower has a stiffness with respect to relative deflections from axial and shear forces and bending and torsional moments. This stiffness is a function of several variables, including the geometric configuration, the mechanical properties and the sizes of the individual members.

Each guy cable also has a stiffness with respect to movement of its attachment point to the tower that is a function of the amount of initial tension, the magnitude of ice load, and the magnitude and direction of wind load on the cables. By evaluating all of these, it is possible to simulate all the guys at a given level as a spring having a specific stiffness. Because of the nonlinearity of some of the relationships involved, the spring constant derived is only valid for a specific set of conditions and for a finite range of translation. Similarly, a torsional spring constant can be derived. It is interdependent with the translation stiffness and is also valid for only a finite range of translation.

Another difference between a guyed and self-supporting tower is the magnitude and significance of the axial load. For a self-supporting tower this is composed only of the gravity loads from the tower, its appurtenances and any ice load. It is independent of wind load, and its effects on individual member loads are relatively small. The axial load for a guyed tower includes in addition to the gravity loads, the vertical components of the tensions in the various guys. Since these tensions are directly affected by the wind load, the axial load is now dependent upon wind load and its effects on the individual leg members are relatively large. Tension in the guy wires also produces an additional bending moment on the tower equal to the product of the axial load and the deflection of the tower.

Despite the complexity of the relationships involved, the availability and widespread use of computer systems permits accurate structural analysis of guyed towers. There are several different structural models that may be used.

One of the most commonly used idealizes the tower shaft as a continuous beam-column on nonlinear elastic supports (the guys) subjected to simultaneous transverse (wind and or seismic) and axial (dead, ice and vertical components of guy tensions) loads.

The modes of failure are buckling of individual leg or bracing members under compressive loads; rupture of bracing members, guys, or guy anchor arms under tensile loading; and shear or tension failures of the connections.

Dynamic Considerations

As previously mentioned, even though wind and earthquakes involve kinetic energy, their effects are simulated by equivalent static loads determined in accordance with the design standards. In recent years there have been more sophisticated efforts to investigate the actual response of tower structures to the dynamic aspects of wind gusts. A conclusion drawn from these studies is that the bending moments in the upper portions of tall, guyed towers are considerably higher than those determined by the usual static analysis. Consequently, the loads imposed on the vertical legs and their splice connections would be amplified beyond safe limits indicating a potential failure condition. Considering the usual fundamental periods of tall guyed towers, it appears that towers taller than 1,200 to 1,300 ft. should be investigated dynamically as well as statically.

There are two other phenomena related to the dynamics of wind that are important in guyed tower design. These are aeolian vibrations and *galloping*, both of which involve periodic loading.

Aeolian vibrations are low amplitude, high frequency movements, which occur in the tower guy cables due to a phenomenon known as vortex shedding. If they are not suppressed through the use of dampers, they can result in destruction of the filaments in the tower lights at the least, or fatigue failure of guy cable and collapse of the tower at the worst. Dampers attached at one or both ends of the guy cables have proven effective in controlling these vibrations and should be considered for all tall, guyed towers.

Galloping is a condition of instability involving large amplitude, low frequency movements. It is caused by the perpetual amplifications of periodic loads due to the motion of the body itself. The most dramatic and well-known example of galloping is the collapse of the Tacoma Narrows suspension bridge in 1940.

For tower structures, galloping is usually associated with the guy cables on tall towers, but in at least one instance it was related to a large rectangular waveguide. There have been several different methods involving detuning and energy dissipation used for preventing galloping in guy cables that appear to be successful. In the case of the rectangular waveguide, galloping was controlled by moving the waveguide inside the tower along the centroidal axis from its original position on the outside of one face. This reduced the torsional rotation of the structure, which was the source of the perpetuating force. Based on this experience, it would appear prudent to always install this type of waveguide inside the tower unless adequate torsional rigidity is provided throughout the height of the tower.

Allowable Stresses & Safety Factors

Towers, like all other structures, are designed so that the maximum anticipated stresses are less than those which would cause failure. This ratio of failure stress to maximum allowable stress is known as the safety factor. It is intended to provide for several variations from the ideal conditions assumed for design including loads greater than anticipated, imperfections in materials and tolerances in fabrication and construction.

The ANSI/TIA/EIA-222-F-1996 Standard refers to the American Institute of Steel Construction (AISC) *Specification for the Design, Fabrication and Erection of Structural Steel for Buildings* for the design of the structure's members and to the American Concrete Institute (ACI) *Building Code Requirement for Reinforced Concrete Structures* for the design of the reinforced concrete foundations and guy anchors.

For towers under 700 ft in height the allowable stresses given in the AISC specification may be increased by one third. For towers 1,200 ft or taller, no increase is permitted. For towers between 700 and 1,200 ft, the amount of increase permitted is determined by linear interpolation.

In a similar manner, the required reinforced concrete strength for towers under 700 ft in height must equal 1.3 times the calculated reactions and for towers 1,200 ft and taller, 1.7 times the calculated reactions. Linear interpolation between these two values is used for towers between 700 and 1,200 ft to determine the required strength.

The minimum safety factor for guy cables is 2 for towers under 700 ft in height and 2.5 for towers 1,200 ft and taller with linear interpolation applied between these two heights.

EFFECTS OF ANTENNAS AND TRANSMISSION LINES

Except for AM radiators, the tower is the necessary evil to support the broadcast antennas and transmission lines at a suitable height above ground. Thus the effects of this equipment are of paramount importance.

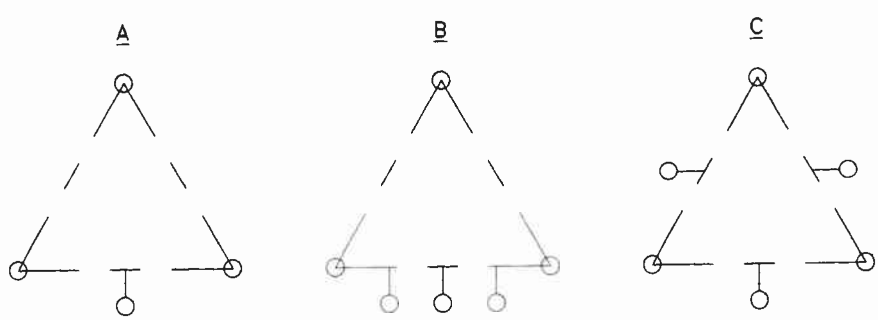
Loads

Every antenna imposes a wind load and a dead load on the tower. If the antenna is mounted atop the tower, it also imposes an overturning moment. If it is mounted on a side of the tower, the antenna imposes a torsional moment. For TV and FM broadcast and microwave antennas, these loads are relatively large, and their location has a significant effect on the placement of guy cables.

Transmission lines feeding the various antennas also impose wind and dead loads on the tower. These loads are distributed uniformly between the antenna and their entry point near the base of the tower. The total produced by a coaxial line or waveguide is frequently greater than that produced by the antenna itself. The shape of the transmission line influences the magnitude of the wind load, with circular or elliptical lines having loads that are 60% of those for rectangular lines with the same projected area.

It is important not to overlook the support system required for transmission lines. Some large waveguides have support systems that require nearly continuous vertical structural members that add substantial wind and dead loads. Small, flexible lines require supports at a maximum interval of 3 to 4 ft, which is often less than the vertical spacing of horizontal members in the tower. Thus, it may be necessary to provide an additional support structure for these lines, again adding to the total load.

An important consideration when locating transmission lines is that the ANSI/TIA/EIA-222-F-1996 Standard permits linear appurtenances that are attached to the tower face, and do not extend in width beyond the normal projected area of the face, to be treated as structural components rather than individual appurtenances for purposes for purposes of calculating wind load. The effect of this is to substantially reduce the magnitude of the calculated wind load on these appurtenances (see Figure 2.2-7). When this procedure is used, it is mandatory that the transmission lines be installed as specified in the design. This may not be



	A	B	C	
RS-222-C	1.00 (Ref.)	3.00	3.00	CYLINDRICAL
222-F	1.53	4.59	1.78	
RS-222-C	1.00 (Ref.)	3.00	3.00	FLAT
222-F	1.70	5.10	2.07	

Figure 2.2-7. Effects of linear appurtenance locations on calculated wind loads.

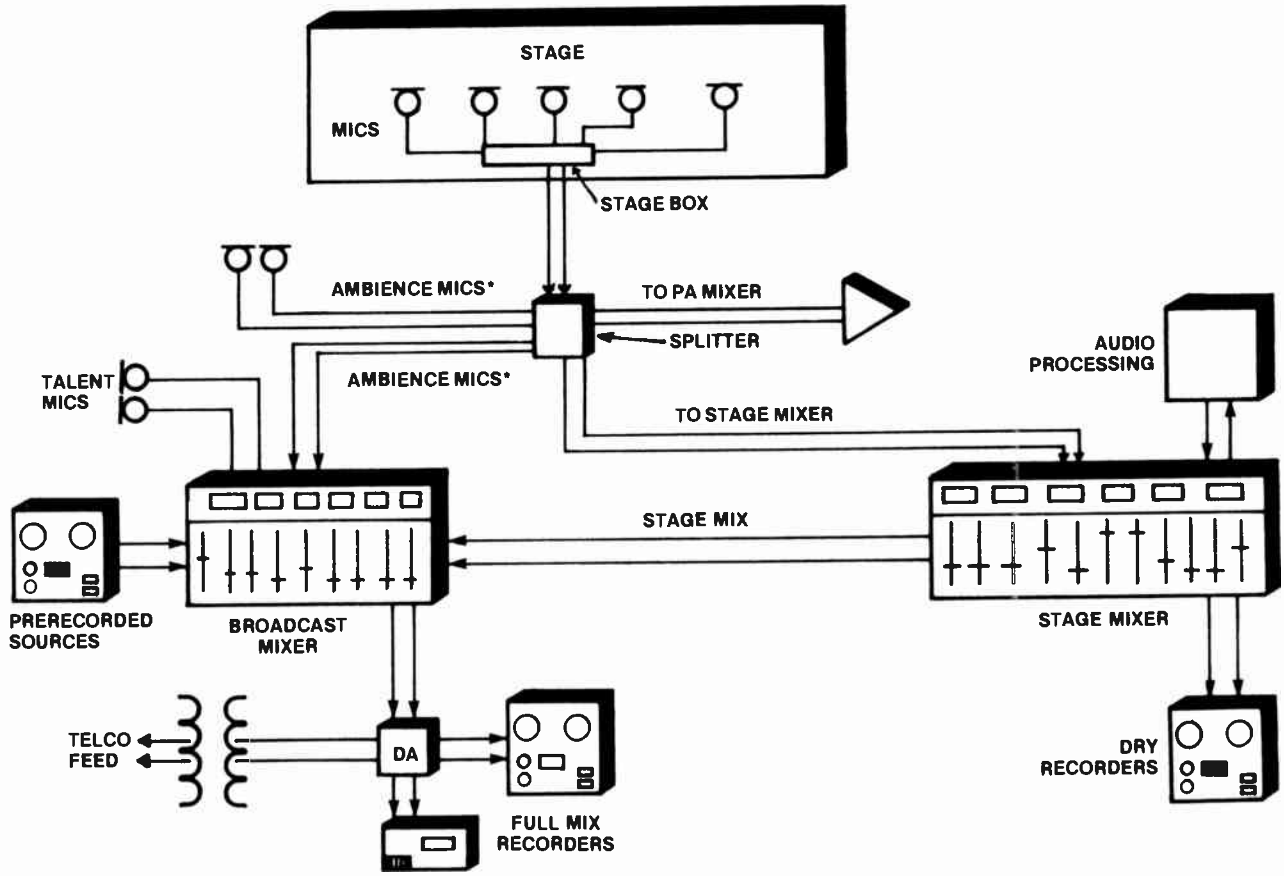


Figure 3.9-14. Typical setup for a large scale remote, using two separate mixing positions. This arrangement is ideal for events such as festivals or conventions, where several separate performances with set-changes in between will be broadcast in a long, continuous program from the site.

remote setup, in which two separate mix positions are used.

When multiple mixers (such as the PA and separate radio truck just mentioned) each require access to the same microphones, a number of options are available. The most primitive would be to put separate microphones up for all mixers, but this becomes extremely impractical and unsightly, especially in the case of handheld vocal or announce microphones. For this reason, a shared arrangement is best, using proper *microphone splitting* techniques. Good transformer isolated microphone splitters are recommended for this. One mixer remains directly connected to the microphones, while the others are fed by the secondary of a bridging input transformer. The microphones see only one load, and the consoles remain electrically isolated from one another. Typical microphone splitters also provide separate ground lift switches for each channel (or in some cases, a single, ganged switch), by which ground loops between consoles can also be eliminated. If condenser microphones are used, their phantom power can only come from the console receiving the direct feed, since dc supply voltage from the transformer fed consoles will not pass back across the transformers.

A cassette or DAT recorder is often useful at a remote broadcast because it gives added flexibility to the remote crew. The recorder input signal can be taken from an auxiliary output on the audio mixer, allowing interviews or material from the PA system to be mixed and recorded for later use on the air. This is always preferred over relying on a recorder at the station end, because the backhaul system may degrade the audio quality, especially when a POTS line is used, but also because the line may fail during the broadcast. In either case, an on site deck provides a good quality recording for later production and broadcast.

In cases where a separate stage mixer or truck is used for the stage premix, it is this *dry* premix that is usually recorded on site. These production recordings therefore include only the mix of the actual event, and do not contain any continuity or other local production elements added at the broadcast mixer. The full broadcast mix can be recorded for archival purposes either at the remote site or at the station.

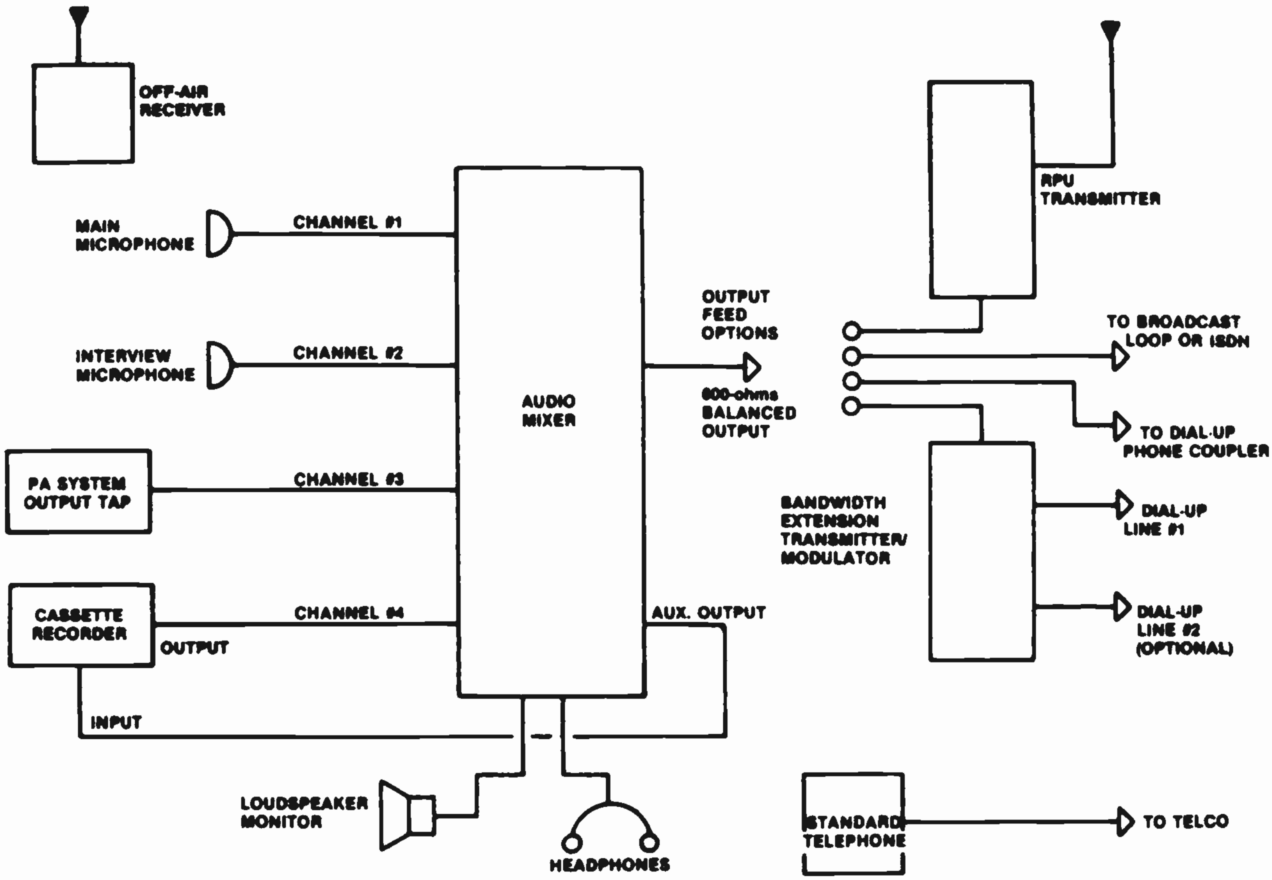


Figure 3.9-13. A typical equipment configuration for a medium scale RENG broadcast.

components below about 60 Hz. At the studio, the receiver audio output would run through another high pass filter to remove any control tone signals from the automation system program channel feed.

A more recent solution available in most cases is the use of the cellular telephone system for this purpose. A portable or transportable cellular phone has become a common part of most RENG systems today. When more than just simple voice communication is required—such as *interruptible foldback* (IFB) or other audio signals to talent headphone or monitor speakers—an audio interface to and from the cellular phone may be used, on both ends of the link. Unlike the standard dial-up system, the cellular phone system is 4-wire end-to-end (it uses a *pair* of RF channels, one for transmit and one for receive), so that no hybrid or gating circuitry is required. The same type of hardware may be used for feeding RENG program audio from the field when no other method is possible. Audio performance will typically be somewhat degraded from what is expected from the standard dial-up phone feed.

REMOTE PRODUCTION EQUIPMENT

The audio equipment used at the site of the remote will vary as widely as the types of programs broadcast. Figure 3.9-13 illustrates a typical production setup used with either a wired or wireless backhaul system.

In this example, a 4-channel audio mixer is used to mix the sources and drive the telco service or RPU transmitter. Careful attention should be given to the connection of the mixer output to a POTS line. A phone coupler should be used between the mixer and the telephone unless the mixer is specifically designed to work directly into a hot dial-up line (one with dc voltage across it). This caution applies to a connection made either to the phone-line tip and ring wire or to the telephone set through the handset terminals.

As shown in Figure 3.9-13, two microphones are used—one for the announcer and another for interview guests. An output is taken from the local PA system to pick up audio from meetings, speeches, music or whatever. For more complex and high budget music remotes, a separate mixer or audio truck is used to *premix* the stage microphones down to a stereo feed, which is then fed to the mixer in Figure 3.9-13. Sound reinforcement (PA) feeds of musical acts are not always inclusive of all the elements required—or are improperly balanced—for the radio mix. For this reason, a separate radio mix is often required if proper radiophonic balance and control of the stage event's mix is critical. Figure 3.9-14 shows a more complex

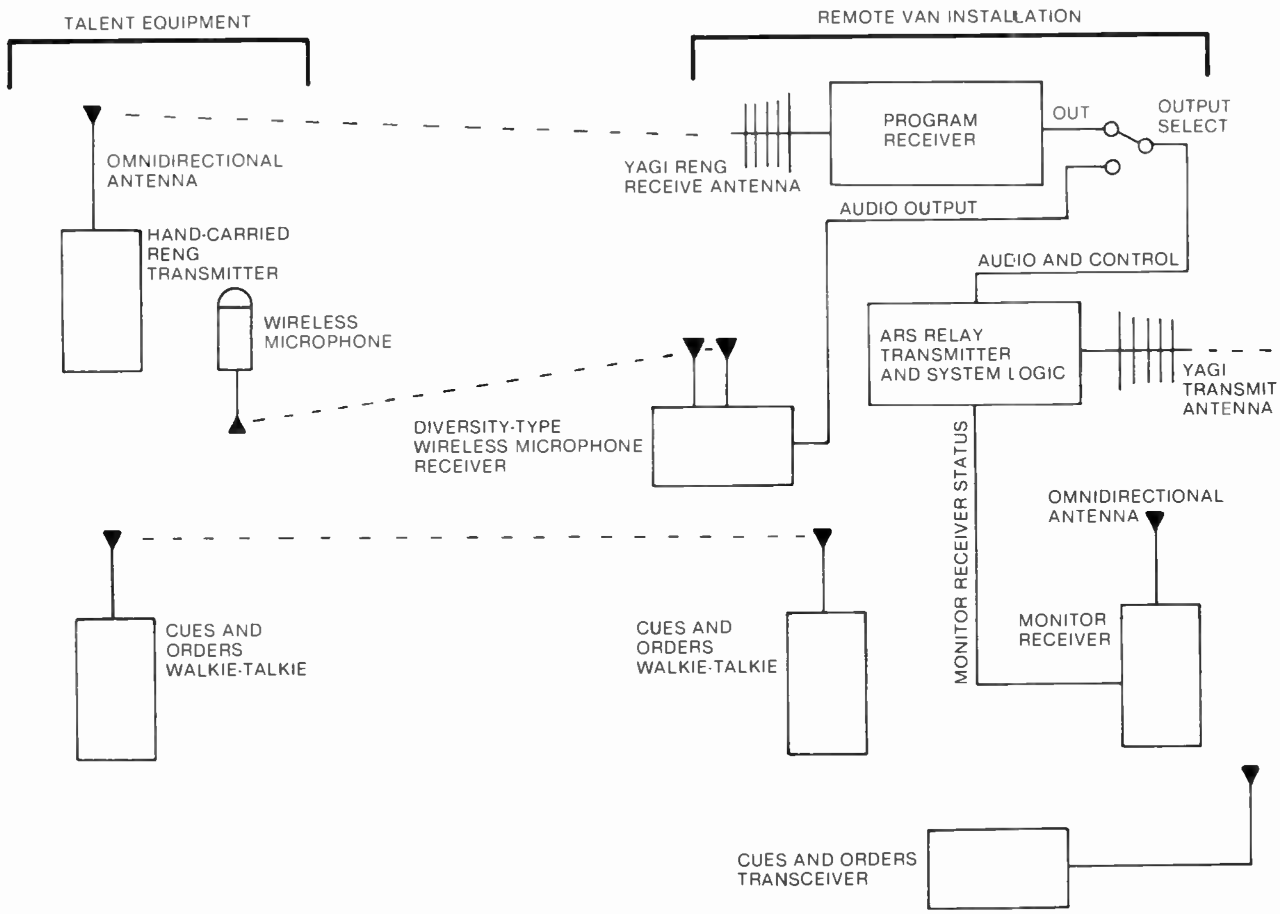


Figure 3.9-12. The use of an ARS system at the event site for added range and talent flexibility.

plest method is an over-the-air cue in which the talent simply listens to the station's air signal and takes cues from the studio announcer or a pre-recorded introduction. Other methods include use of an FM station's subcarrier signal for cueing information or a separate, dedicated, radio link specifically used for cueing instructions, either from the remote truck (as shown in Figure 3.9-12) or from the main studio.

If a station needs a more sophisticated intercommunication system, a trunked 800 MHz radio system can be considered. A 5- or 10-channel trunked repeater acts like a small telephone exchange in which the number of users (telephones) exceeds the number of channels (trunk lines). Telephone system theory is used to predict the busy level that can be expected during periods of heavy radio traffic. Three minute time out timers are usually included in mobile transmitters to enforce time limits.

These trunked systems can tie into the regular telephone system at hilltop repeater sites or at trunked base stations. Broadcasters interested in 800 MHz trunked radio should contact their local area land mobile operator to see if such a system is available.

In certain situations, a station may be able to design and license a UHF business radio system for dispatch and coordination of RENG crews. These systems offer the user the luxury of not encountering a busy signal, as may occasionally happen in a trunked system. As with the trunked network, no programming is allowed on a UHF business radio system.

One of the problems often encountered when carrying remote broadcasts on an automated station is the need to have an operator standby during the broadcast to trip the automation system to the next event when the talent at the remote site gives the proper cue. The simplest way around this problem is through the use of a sub-audible tone that is high enough in frequency to not interfere with the ARS sub-audible tone that may be used for repeater equipment, and low enough in frequency so that it does not interfere with normal program audio. For example, in a system where the ARS access sub-audible tone is 25 Hz, an *advance system* control tone of 45 Hz could be used. In order to prevent premature automation system trip commands, the program audio input to the remote location transmitter would be passed through a high pass filter to remove any audio

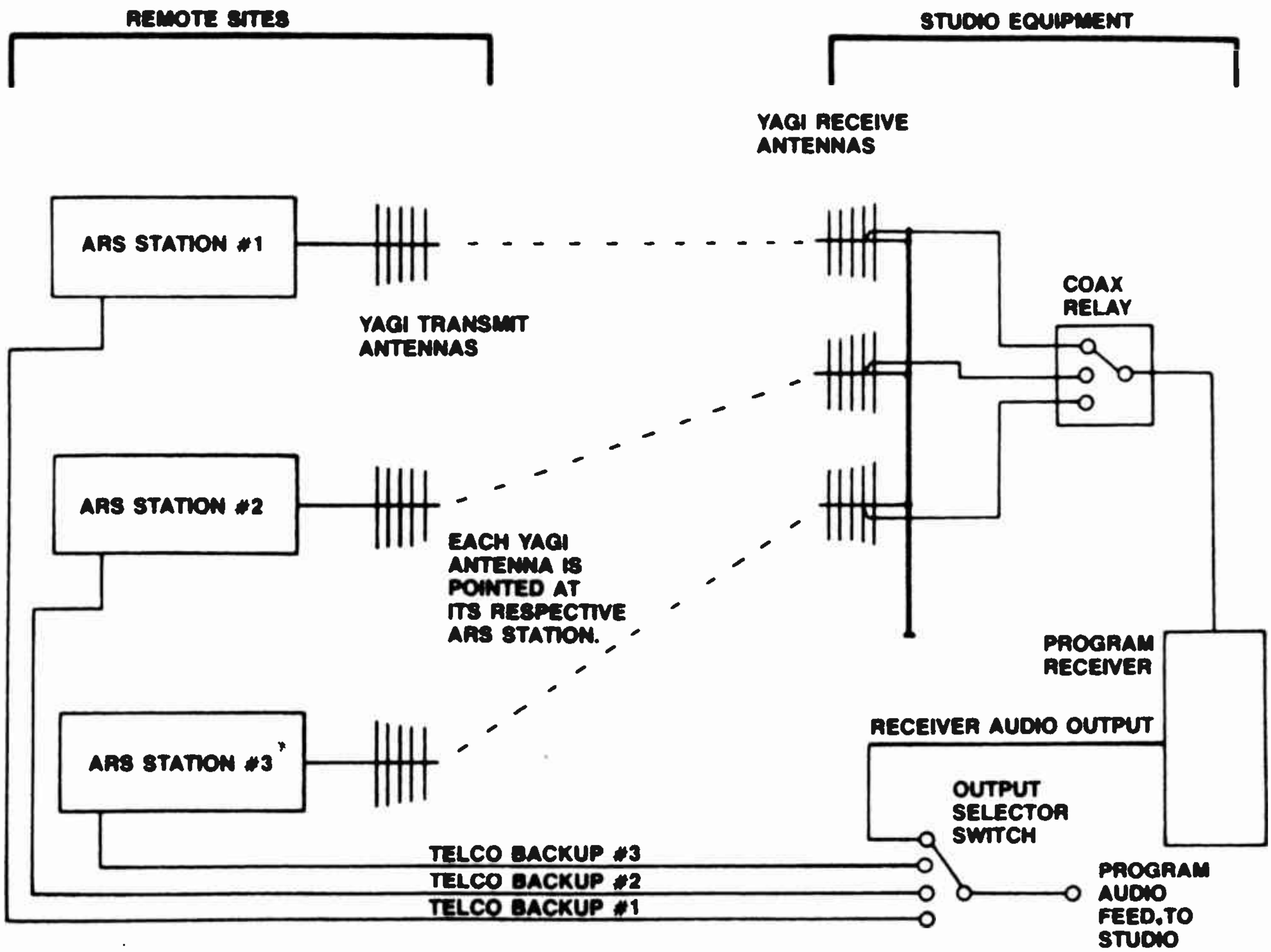


Figure 3.9-11. A multiple site ARS network feeding a central studio control point. Note each ARS station is also connected to the studio via a land line for backup protection.

allows the receiver to squelch and generate a loss-of-carrier command to the ARS system logic. In order to maintain positive control over the system, a means should be provided to override the ARS logic by remote control from the studio.

Figure 3.9-12 shows some of the ways the remote location program audio can be transmitted. As mentioned previously, the communications transceiver is used for cues and orders from the studio location. The program channel signal can consist of a hand- or pack-carried transmitter, which directly feeds the studio receiver or one or more ARS systems. Figure 3.9-12 also shows a repeater station configuration that can be used when a high power transmitter is required to reach either the studio or the ARS relay point. The use of a repeater—configured as a standard ARS station—in a car or van outside the remote location also gives the talent at the event greater flexibility, because a small hand-carried transmitter can be used there, rather than a larger unit with antenna and power cables attached. This arrangement is also ideally suited for use with a wireless microphone, which gives the talent an even greater degree of flexibility. The receive antenna at the remote van can be either an omnidirectional unit or a Yagi. The system shown in Figure 3.9-12 includes a monitor receiver to prevent ARS transmission over traffic already in progress.

There is a limit, of course, to the number of times a signal can be repeated and still maintain good audio specifications. Moreover, each added hop in the path between the remote site and the studio increases the chances of a spurious signal interrupting the remote feed. Each additional hop also increases the complexity of the system and the vulnerability of the link to equipment failure. The design goal for any RENG system should be to keep the arrangement as simple and direct as possible, while still providing talent flexibility, backup protection and high performance.

Remote Cues and Orders

Communications with a remote crew from the studio can be accomplished in one of several ways. The sim-

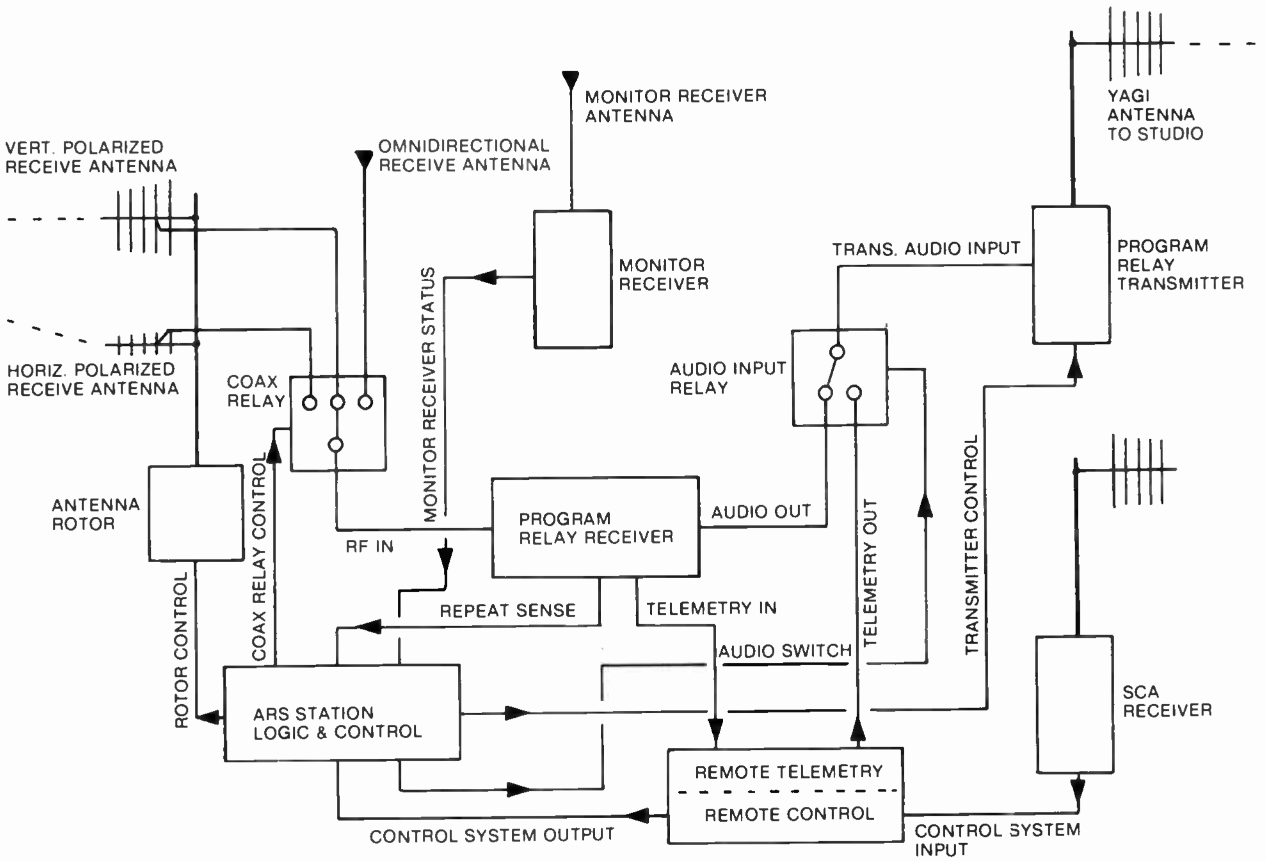


Figure 3.9-10. A high performance secure channel ARS station with remote control of system functions.

necessary switching and positioning work at the ARS site. For stations with multiple site capability on the main transmitter remote control system, the ARS remote points can be simply treated as other *transmitter sites* and controlled as such from the master unit.

A monitor receiver is included at each ARS installation to inhibit activation of the ARS transmitter if a transmission is already in progress on that frequency. As shown in Figure 3.9-10, the control commands are received over a subcarrier receiver from the main station transmitter. The relay station logic interfaces the remote control unit with the receive antenna coaxial switch and the antenna rotor control box. During set up, the telemetry section of the remote control unit provides an audio FSK signal that is sent back to the studio control unit via a telco line or the ARS transmitter, as shown.

The SCA, omnidirectional program, monitor and relay transmit antennas are all fixed in position. Only the directional receive antennas, one set for horizontal polarization and the other for vertical polarization, are movable. An arrangement such as that shown in Figure 10, will provide maximum flexibility and minimum risk of program audio disruption.

In a system where two or more of the ARS stations shown in Figure 3.9-10 are used, an arrangement such as that shown in Figure 3.9-11 may be implemented. The studio remote control unit is used to determine which of the ARS stations is allowed to repeat the program traffic. Those stations that will not be used to repeat the program material would be instructed by the studio operator to remain inactive. For multiple site ARS operation, as shown in Figure 3.9-11, individual directional receive antennas, or a single directional receive antenna mounted on an antenna rotor, may be used to receive the ARS traffic at the studio.

For protection against system failure, an equalized telco loop can be installed between each ARS point and the main studio. With this backup provision, an equipment failure in the relay gear would not interrupt a remote broadcast.

One of the problems sometimes experienced with ARS equipment is the possibility of a desired signal opening the system, and an undesired signal keeping it open after the desired traffic has ended. This can occur if the tone burst method of repeater keying is used. For example, a valid tone burst signal unlocks the ARS system and then undesired noise or traffic holds the channel open after the desired traffic has ended by prohibiting a loss-of-carrier indication from the receiver. The ARS will thus be stuck open until the level of the interfering signal drops to a point that

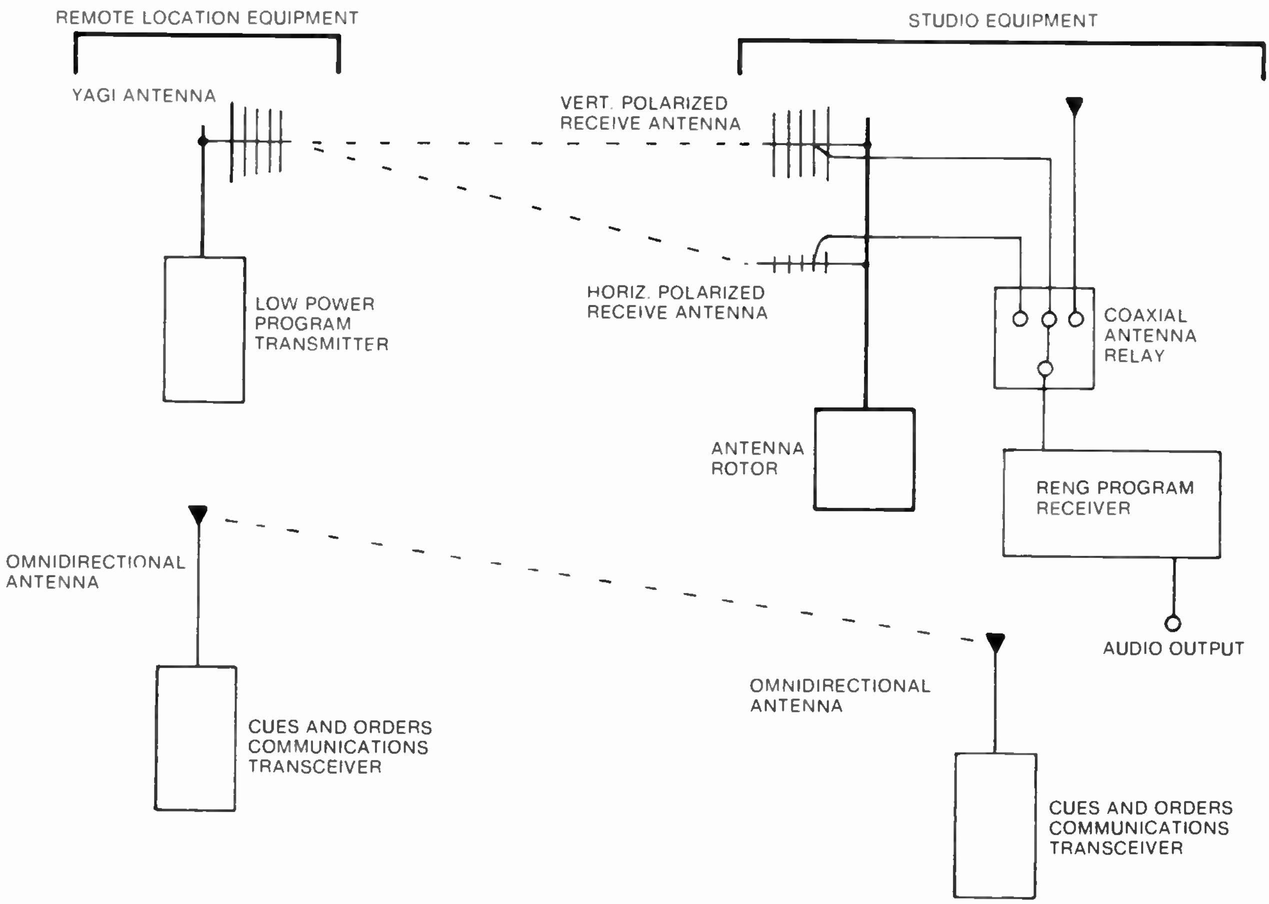


Figure 3.9-9. A high performance 2-point RPU system designed for operation in frequency-congested areas.

requiring wide frequency response and a high signal-to-noise ratio. The lower power program relay transmitter and directional receive and transmit antennas provide a secure and quiet channel, without causing interference to other RPU band users in the area.

Between remote broadcasts and when beginning the initial set up procedure for a remote, the omnidirectional antenna is patched into the broadcast quality RPU band receiver at the studio through the coaxial antenna relay. Once contact has been established with the remote crew, one of the directional antennas—which are mounted on a common mast driven by a remote-controlled antenna rotor—is switched into the studio receiver. The polarization of the transmission from the remote site is planned before the remote crew leaves the studio. Selection of either horizontal or vertical polarization is made during the frequency coordination process, or at the discretion of the user. Engineers may find that a particular polarization may yield better results from certain geographical areas, and in such cases, that polarization would be chosen. Once the proper antenna has been selected, the antenna rotor is adjusted for maximum received signal strength. The studio operator then talks the remote crew into the best position for its Yagi transmit antenna. At this point, the antennas are locked down and the link is ready for the remote broadcast.

If a variable power transmitter is used at the remote site, or transmitters of various power levels are available, the transmitter power would next be adjusted to the point necessary to achieve the required S/N performance at the receiver. After power output adjustment, the antennas on both the receive and transmit ends should be checked again for correct positioning.

While this process may be time consuming and require the purchase of additional equipment, it will assure a high quality, secure RF link from the field to the studio. This system will also result in a minimum of unwanted radiation to other RPU band users.

Figure 3.9-10 shows a high performance, secure channel ARS. The same antenna selection and positioning procedure is used in the ARS installation as was used in the 2-point system of Figure 9, except that the antenna switching and positioning work is done by remote control. The link for this remote control system can be a subcarrier on the main station broadcast signal, a separate dedicated radio link, a dial-up telephone patch or a leased telco data or voice loop. A standard broadcast transmitter remote control system is used, with the common channel on-off/up-down functions performing the is side mounted on the station's broadcast tower, and the RPU receiver system is housed in the shack or transmitter building with the station's broadcast transmission equipment. RPU audio is routed from the receiver output back to the studio via a high quality transmitter-to-studio link (TSL). Simple control of the RPU receiver (power on/off, frequency selection) can be broken out on spare relays or closures from the broadcast transmitter's remote control system.

A more sophisticated approach may be called for in rougher terrains. The system configuration shown in Figure 3.9-8 overcomes the geographical coverage area problem through the use of an automatic relay station (ARS). The range of a RENG system can be greatly extended through the use of an ARS. Such systems also make it possible to use lower power transmitters in the field, since the transmitter at the program origination point need only be powerful enough to reach the ARS site. This often allows the use of smaller and lighter remote transmitters, usually hand-carried or pack-carried units. The arrangement shown in Figure 3.9-8 will satisfy the requirements for wide area coverage and is sufficient for radio markets where spectrum congestion is not a problem. Because all antennas in the system are omnidirectional, however, the configuration is not suitable for use in larger urban areas which are experiencing frequency allocation problems. For such applications, a more sophisticated approach is needed to RENG activity.

Figure 3.9-9 shows a high performance 2-point RENG system designed for operation in spectrum congested areas. At the remote site, two transmitters and two antennas are used. The communications transceiver is used for conveying setup information between technical personnel at either end, and to relay cues and coordinating information. The low power transmitter and its associated directional transmit antenna are used to relay the program audio signal to the studio. At the studio site, a communications transceiver, feeding an omnidirectional antenna, is used for the setup information, cues and coordination work. The multi-antenna receive system is used for program audio pickup.

The *cues and orders* radio system shown in Figure 3.9-9 is used for general purpose communications not

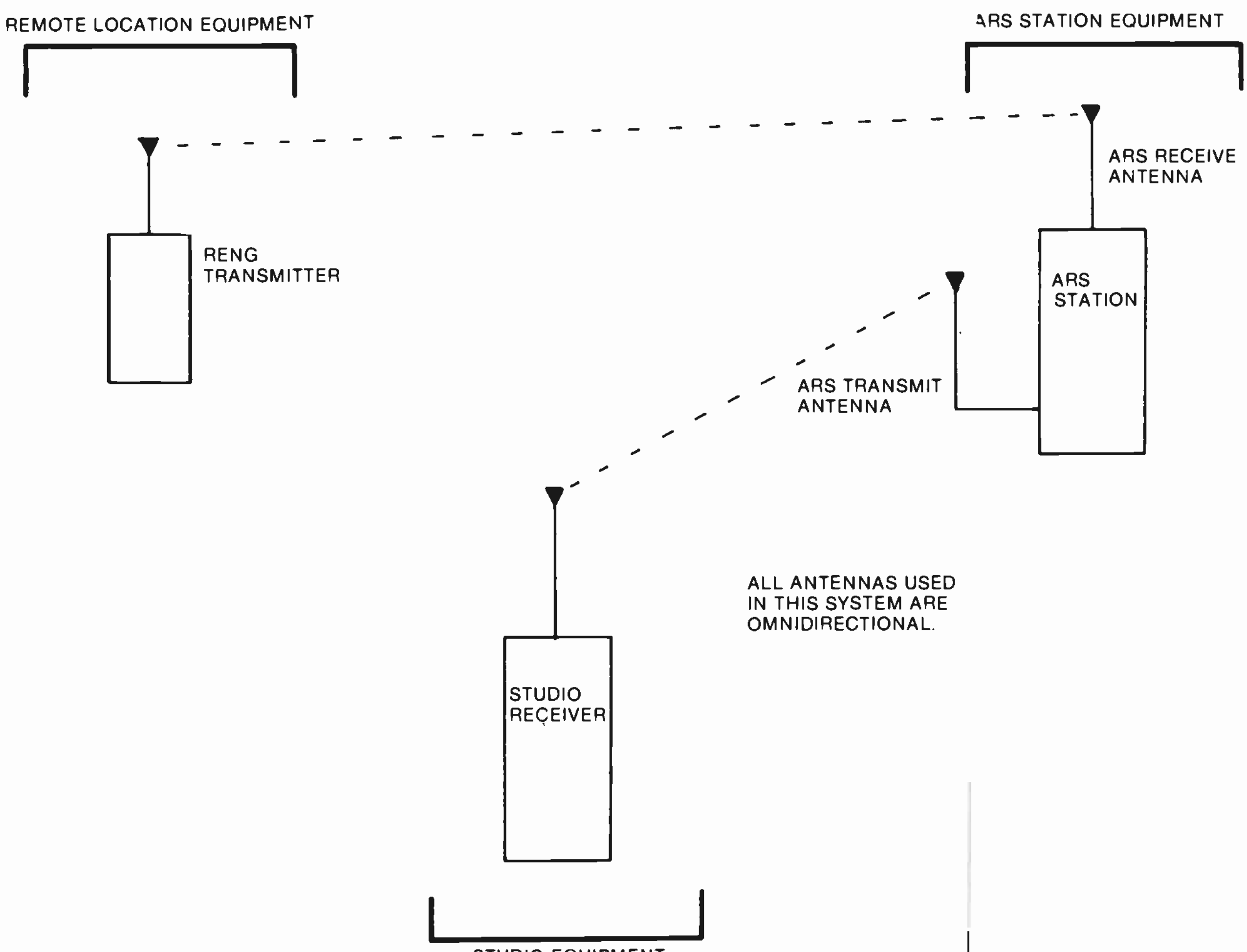


Figure 3.9-8. The basic RENG program relay configuration using an ARS station between the remote location and the studio.

of medium level co-channel signals. Preselectors are often necessary at mountaintop or antenna farm locations because of the high level RF signals present at such sites. It is not uncommon to have a 1 kW land mobile paging transmitter operating in the 454–455 MHz range located near an RPU band receiver that is working in the 455–456 MHz frequencies. High power FM or TV transmitters can also cause desensitizing of the receiver front end, unless adequate bandpass filtering has been included in the receiver design.

The locations commonly used for relay sites are seldom ideal from an environmental standpoint. They are often inaccessible during portions of the year, very hot in the summer and very cold in the winter. For these reasons, rugged equipment should be selected, in order to minimize downtime. Temperature extremes can also cause problems for frequency determining elements, as well as accessories such as cavity filters, preselectors and preamplifiers. Since relay sites are often difficult to reach, equipment should be designed for easy maintenance, preferably through module replacement. A spare stock of modules should be kept at the site so that the system can be quickly returned to operation. The defective module can then be serviced at the studio, or returned to the factory for repair.

Regular performance tests should be made of the RENG system, just as an engineer would do with any other important chain of equipment at the station. Regular checks and measurement often allow the engineer to spot problems that could cause a total system failure if left unattended. If trouble is experienced with a piece of receiving equipment, the possibility of interference from other services should not be overlooked. A spectrum analyzer is invaluable for such work.

System Configuration

The requirements of users will vary greatly from one station to the next and from one market to the next. There are, however, several standard system configurations that can be modified to fit the requirements of most users. These range from the simple point-to-point program relay system common on many small-scale operations, to complicated multi-point relay installations with automatic signal-quality voting circuits.

Figure 3.9-7 shows the basic RENG program relay system in which one (or more) transmitters on a particular frequency are used in the field, and a signal receiver is located at the studio. All antennas used in the system are omnidirectional. While there is much to be said for system simplicity, such an arrangement is not practical in an increasing number of urban areas because of spectrum congestion problems and the need to cover large geographical areas.

A simple alternative approach for a station with a good yet conveniently located transmitter site (not co-located with the studio) is as follows. The RPU antenna

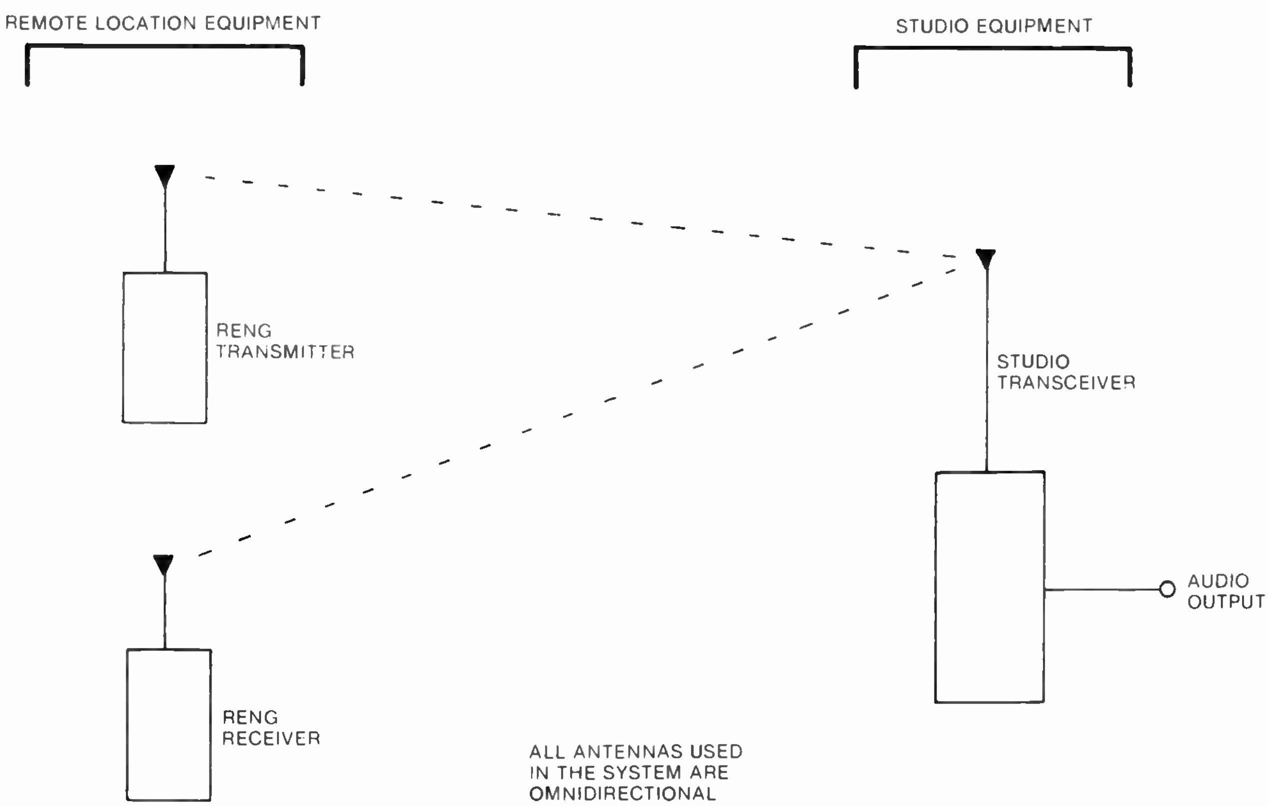


Figure 3.9-7. The basic RENG program relay system using a single hop from the remote location to the studio.

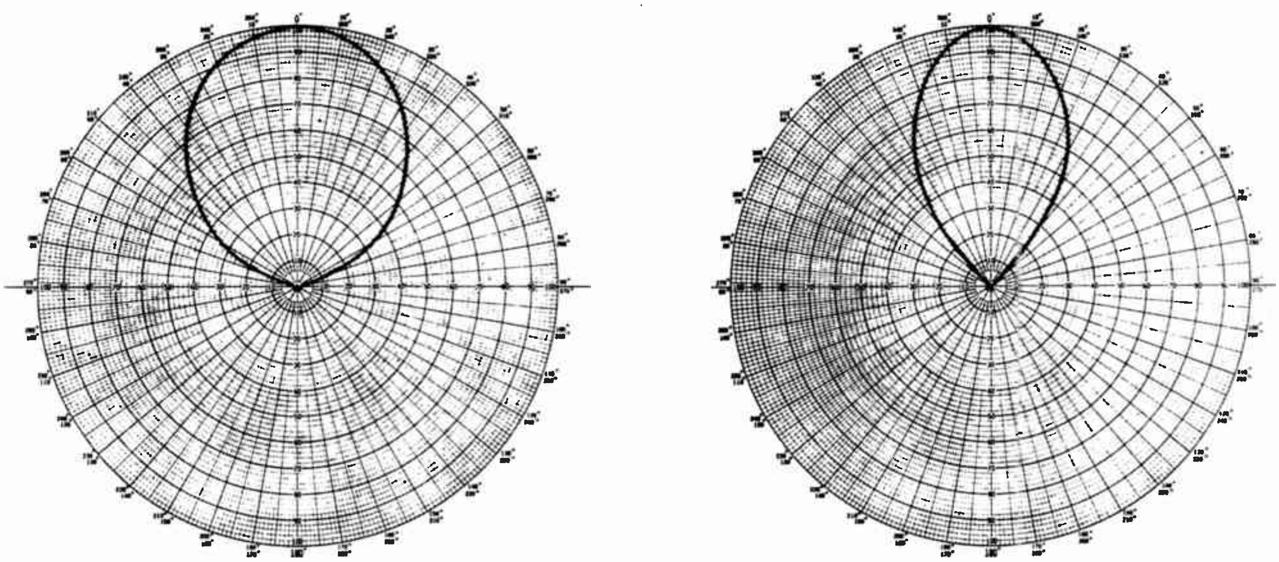


Figure 3.9-6. The radiation patterns for the Scala CL-400 broadband log periodic antenna, designed for use in the 450 MHz RPU frequency band. (Courtesy of Scala Electronics.)

back ratio of approximately 14 to 18 dB. Figure 3.9-5 shows the radiation pattern for a commonly used 5 element 150 MHz Yagi. This particular antenna measures 40″ × 40″ × 4″ and weighs 8 lbs. Thus it is small and light enough to be used on remote broadcasts. It is also suitable for permanent installations using either horizontal or vertical polarization. These antennas may be stacked in 2 and 4 bay arrays (with suitable phasing harnesses) for additional gain and directivity.

Most Yagi antennas are made to match the specific frequency requirements of the user. Multiple frequency operation using a single antenna is possible, however, with reasonable VSWR numbers, as long as the operating frequencies are not removed from the cut center frequency by more than 1 to 2%.

A more recent addition to the RENG user's bag of electronic tricks is the broadband log periodic antenna, which can be used on any channel within a wide band of frequencies. Such antennas provide a smooth pattern with minimal sidelobe radiation and a high front-to-back ratio (typically 25 dB in the 150 MHz band). Nominal gain for 150 MHz operation is 7 dB. Units can also be stacked to provide additional gain and directivity. Such antennas are usually larger and heavier than the familiar Yagi; however, they allow use of the antenna for virtually any frequency within the specified band at low VSWR levels (a maximum of 1.5-to-1 is typical). Figure 3.9-6 shows the radiation pattern of a log periodic antenna designed for use in the 450 MHz band. Horizontal or vertical polarization is available. The antenna shown in Figure 3.9-6 has a gain of 8 dB and a front-to-back ratio of 35 dB. Such a unit is therefore well-suited for operation in areas with high spectrum congestion.

Just as a TV or FM broadcast antenna must be protected against icing problems, so should antennas used in RENG applications. Although antenna deicers are not used in RENG installations, a radome is often available for an antenna to protect it from damage or degradation in performance due to snow, ice or salt spray.

Transmitter-receiver Considerations

Whatever the configuration of the planned RENG system, there are several important points that should be considered. Most of these items apply to receiving equipment, which usually present the greatest problems to a system designer. Transmitting equipment must also be selected with care, but the receiving links in a RENG system are the ones most often subjected to conditions that may make good performance difficult.

A receiver that has sufficient dynamic range and headroom should be selected to allow the system to deal with strong adjacent-channel signals, as well as very weak and very strong co-channel signals from transmitters in the network. A receiver with inadequate headroom will clip and yield distortion. Wide dynamic range active devices should be used in the receiver front-end, such as *gallium arsenide field effect transistors* (GAsFETs).

The need for a preamplifier or cavity preselector network ahead of the first RF stage should also be considered. RF preamplifiers can add sensitivity, but they can also cause overload conditions in the presence

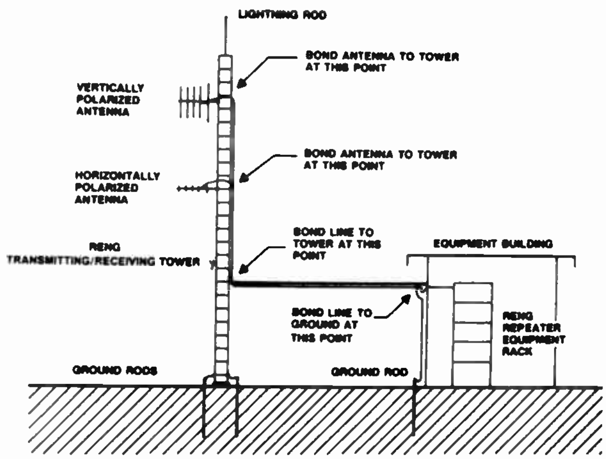


Figure 3.9-4. The recommended installation practices for RENG antennas and transmission lines.

wrapped with good quality tape. Unless this is done, rain may eventually work its way into the connector and cause signal loss or VSWR problems. The line should be grounded (using a recommended grounding kit) at the point where it leaves (or enters) the equipment building and where it starts its climb up the tower (unless the vertical distance to the antenna is less than 10 ft.). This will prevent any high voltage transients caused by lightning from entering the equipment building, and thus the RENG equipment. The advantages of using a low loss line are illustrated in Table 3.9-1. The two ends of the transmission line (at the receiver and transmitter) are probably the easiest parts of the hardwired system in which loss can be introduced, and so care should be taken to install the lines and connectors according to good engineering practice.

A short length of flexible coax is generally used on each end of the two transmission lines for connection to the equipment and antennas (when 1/2 in. or larger coax is used). This *pigtail* is normally no more than 18 in. long (see Figure 3.9-4).

Antenna Considerations

The selection of an antenna for use in a RENG system is an important decision because of the effect the antenna has on system performance and spectrum usage. The usual RENG antenna has, until recently, been the omnidirectional vertical whip with a small amount of gain. Many system planners, however, are now being forced by interference concerns to use directional antennas with moderate amounts of gain. The low power levels commonly used with RENG equipment and the RPU band frequencies make it possible economically to achieve increased ERP through the use of high gain transmit antennas. The use of high gain antennas also concentrates the radiated signal where it will do the most good, and minimizes radiation in directions that may adversely affect the RENG activities of other stations in the same, or nearby, communities.

The omnidirectional base station antennas commonly used in the 150 MHz and 450 MHz bands are vertically polarized units with 4 to 6 dB gain. Electrical beam tilt is sometimes available. Depending upon the manufacturer, up to 20° downtilt can be provided on 150 MHz antennas, and up to 11° is common for 450 MHz omnidirectional units. Large amounts of beam tilt are normally used when the antenna is to be mounted on a structure that is substantially above the surrounding terrain, thereby improving the antenna's close in coverage.

The typical directional RENG antenna is a medium gain 5 element Yagi. Such a unit provides about 9 to 10 dB gain over a reference dipole, with a front-to-

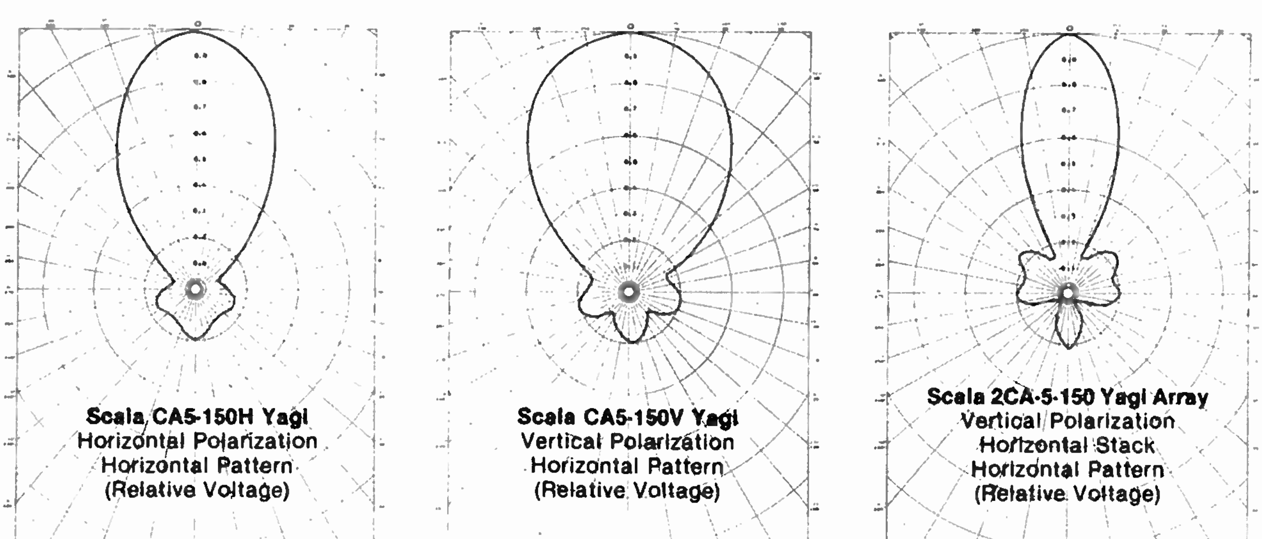


Figure 3.9-5. Radiation patterns for the CA5-150 5-element Yagi antenna made by Scala Electronics for use in the 150 MHz frequency band.

Now, the fade margin can be calculated:

Fade Margin (dB) = $G_s - L_s - R_m$

Where: G_s = total system gain (dB)
L_s = total system losses (dB)
R_m = minimum signal strength required for target S/N (dBm, a negative number)

G_s and L_s are determined by the equations previously shown. R_m (receiver sensitivity) is determined from the receiver manufacturer's specifications. If the manufacturer gives a receiver sensitivity figure in microvolts, the following formula can be used to convert to dBm:

$$R_m = 20 \log \times (V_r \times 10^{-6}/.7746)$$

Where: R_m = minimum required signal strength (dBm)
V_r = receiver sensitivity (microvolts)

R_m can also be found by using a communications receiver test set, available at most commercial and industrial communications radio service shops.

In order to predict accurately the performance of a RENG radio link, the value of R_m must be determined carefully. Many receiver manufacturers specify V_r for 20 dB of receiver quieting. This level is a convenient reference point; however, it should not be used for fade margin calculations. For maximum system performance and reliability, the fade margin determination should be made based upon the signal level required to provide the minimum acceptable receiver signal-to-noise performance.

The recommended fade margin for a 150 MHz band RPU system is at least 10 dB plus 2 dB for each 10 miles of line-of-sight path distance greater than 10 miles. At 450 MHz, the fade margin should be increased to a minimum of 15 dB plus 3 dB for each 10 miles of path distance greater than 10 miles. These fade margins are designed to limit periods of performance degradation of the radio link to 1% or less during worst case environmental conditions. The fade margin assumes transmit and receive antenna clearance above the ground and all obstructions of 50 to 100 ft.

While it is important to provide an adequate fade margin, needlessly high fade margins should be avoided because of the spectrum congestion problems that may result.

Other Planning Considerations

Path engineering for remote location broadcasts is seldom done for RENG activities because of the transient nature of such events. Rough estimates should be made, however, of the geographical areas of interest before attempting remote feeds. It is well worth the time spent to conduct a coverage survey of the primary areas of interest for RENG activity when planning an overall system to determine which locations provide good or marginal performance. A little planning and work ahead of time will save many problems (and probably dead air) once the system is put into operation.

For base or relay installations that, for one reason or another, cannot use frequencies in the RPU bands for program or control data interconnection, there is often an alternative to the telephone company audio loop back to the studio. Private common carriers in many markets are installing competitive audio circuits, which can be leased. If the RENG installation is co-located with a television relay station, it may also be possible to backhaul the RENG audio on one of the TV microwave system's subcarrier channels. The selection of receiving sites must be made with care, keeping in mind the area of coverage required of the receiver. The best location for a RENG system is not always the highest building in town. Placing a receive antenna at a high elevation in a metropolitan area can result in poor performance of the system in the downtown area, since the gain of many omnidirectional vertically polarized antennas decreases as the antenna is raised above the transmitting point. Tall buildings are excellent for point-to-point relay transmissions, but are generally unsatisfactory for wide area coverage in a metropolitan region.

An inexpensive installation option is available to AM broadcast stations that do not want to erect a separate RENG transmitting tower at the main transmitting site. An isocoupler can be installed at the AM tower base that will pass the RPU transmitter frequency with good efficiency (90% is typical), while at the same time presenting a high impedance to the AM band energy. Isocouplers are available in various frequency and power ranges. Installation of these devices may change the base impedance of the AM tower slightly thus an engineering consultant should be contacted before installation work begins.

Every effort should be made to locate the receiving antennas of a RENG system as far away from high power transmitting antennas as possible. This should be attempted regardless of the frequency separation between the receive unit and the suspect high power transmitting antenna. Failure to achieve adequate physical separation may require the installation of filters of various types on the receiver front-end.

In order to keep system losses to a minimum, a low loss transmission line should be used, such as the 1/2-in. foam filled coax shown in Figure 3.9-3. The transmission line and connectors must be made watertight if exposed to the elements. Each connector should be sealed with a silicone dielectric compound and then

Table 3.9-1
Typical Cable Loss for Popular Types of Transmission Line

	Transmission Line Comparisons					
	150 MHz		450 MHz		950 MHz	
	L	E	L	E	L	E
RG-58/U	6.0dB	25%	12.0dB	6.3%	20.0dB	1%
RG-8/U	2.5dB	55%	5.0dB	31%	9.0dB	13%
LDF4-50A	0.85dB	83%	1.7dB	67%	2.5dB	55%
LDF5-50A	0.48dB	90%	0.9dB	85%	1.55dB	71%
LDF7-50	0.28dB	94%	0.56dB	88%	0.88dB	84%

L = Loss in dB per 100 feet.
E = Approximate power transmission efficiency of 100-ft. length.
Note the poor performance figures for RG-58/U. LDF4-50 is 1/2-in. foam filled line, LDF5-50 is 7/8-in. line and LDF7-50 is 1-5/8-in. line. (Courtesy of Scala Electronics.)

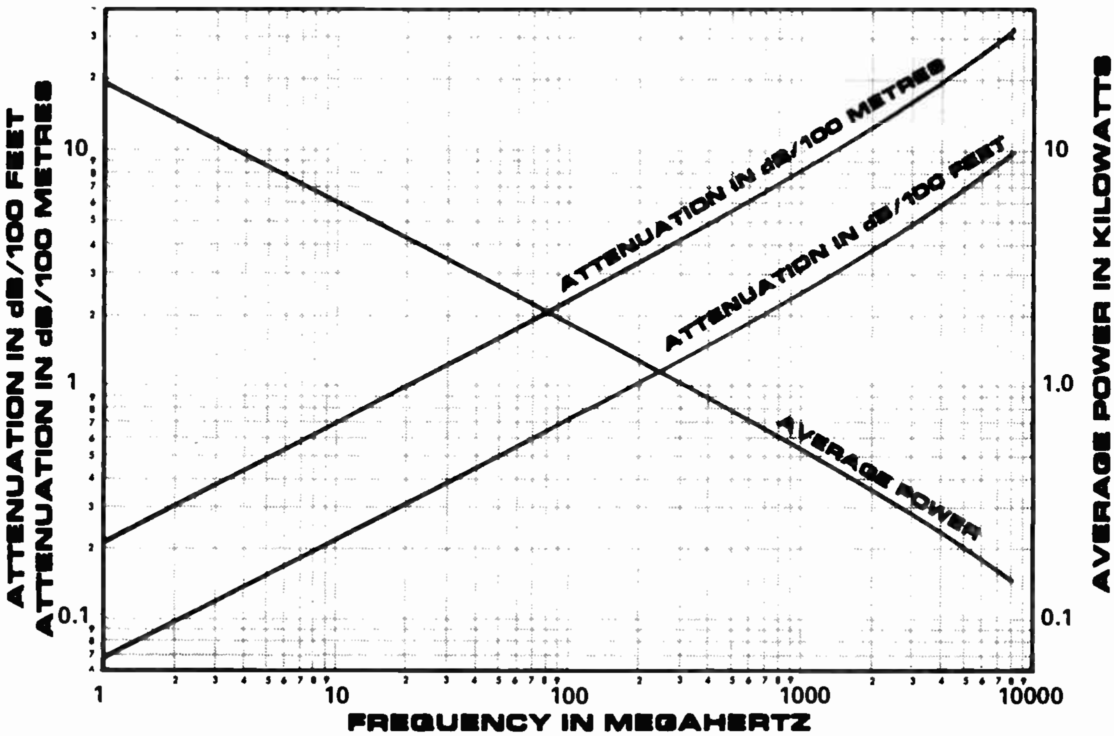


Figure 3.9-3. The attenuation and power handling ratings for 1/2-in. foam dielectric coax. (Courtesy of Andrew Corp.)

change either the transmitting or receiving antenna height or to employ a diversity reception system, if a long term solution is needed.

Determining the Fade Margin

A gain and loss balance sheet should be computed to determine the fade margin of the proposed system. An adequate fade margin is vital to reliable performance of the system, because a link that is operating on the edge of the minimum acceptable receiver quieting will encounter problems later down the road. Normal component aging in the receiver or transmitter can cause a loss in received signal level and thus degrade the system performance. Likewise, new construction near the transmitting or receiving site can degrade the path, resulting in poor performance. Atmospheric conditions—such as severe weather in the area or ice on the transmitting or receiving antennas—can also cause sharp fading, and even a complete loss of signal, if an adequate fade margin above minimum receiver quieting is not provided. The RENG system fade margin can be computed by using the following equations:

$$G_s = G_t + G_{ta} + G_{ra}$$

Where: G_s = total system gain (dB)
G_t = transmitter power output (dBm)
G_{ta} = transmit antenna gain (dBi)
G_{ra} = receive antenna gain (dBi)

The values for G_{ta} and G_{ra} are gathered from the antenna manufacturer's literature.

(Note: dBi = dBd + 1.1 dB, approximately.) The value for G_t is given by the following formula:

$$G_t = 30 + 10 \log P_o$$

Where: G_t = transmitter power output in dBm
P_o = transmitter power output in watts.

Next, the system losses are computed:

$$L_s = L_p + L_l + L_c + L_m$$

Where: L_s« = total system losses (dB)
L_p = path loss (dB)
L_l = transmission line loss (dB)
L_c = connector losses (dB)
L_m = misc. losses (dB)

The values for L_t and L_c can be determined from the manufacturer's literature. Figure 3.9-3 shows typical loss values for 1/2-in. foam filled transmission line. A reasonable value for connector loss with components normally used in 1/2-in. transmission line installations is 0.5 dB. The value for L_p can be found by using the following formula:

$$L_p = 36.6 + 20 \log F + 20 \log D,$$

Where: L_p = free space attenuation loss between two isotropic radiators (dB)
F = Frequency of operation in MHz
D = the distance between the antennas in statute miles

possible, but in the majority of cases, continuously variable power output transmitters are not available. User modification of existing equipment is not an acceptable solution, since such work would most likely invalidate the transmitter's FCC type acceptance.

A more practical solution, therefore, is to purchase RENG transmitters of several different power levels operating on the same frequency (or frequencies). All of the popular RENG broadcast equipment manufacturers offer units with different power output levels. With some equipment, a low power transmitter is used and an optional power amplifier module is added between the transmitter and the antenna to give the needed RF output.

Directional receive and transmit antennas are a good idea from both an efficiency and coordination standpoint. The use of a pair of high gain antennas make it possible to achieve a much greater ERP for the same transmitter power. Of equal benefit in a crowded urban area is the elimination of any non-essential radiation. Through the use of directional transmit and receive antennas, stations can establish more secure channels by placing the radiated energy where it will do the most good (from the transmit end), and rejecting unwanted signals from other directions (at the receive end).

A simple and sometimes effective coordination tool is cross polarization. Two stations on adjacent frequencies may achieve as much as 25 dB RF isolation through the use of different polarizations of transmit antennas, matched by like polarization at their respective receive antennas. Cross polarization results in varying degrees of success, depending upon the frequency of operation and the surrounding terrain. Line-of-sight paths usually will provide good results, but urban centers with their highly reflective buildings, generally cause polarity shifts in the transmitted signal that may significantly reduce the benefits of cross polarization.

Path Engineering for Fixed Stations

Careful path engineering should be performed prior to any licensing work to determine if the proposed locations of base station and ARS installations will be able to achieve the desired results without using excessive amounts of transmitter power. There is much more to path engineering for a RENG system than simply pointing the transmitting and receiving antennas at each other (when directional antennas are used) and turning the equipment on. Base station and ARS systems are fixed position installations that cannot always be located in the best possible geographic locations because of space availability problems, excessive construction or site rental costs, or local or federal licensing difficulties. In such cases, the required path is not the ideal path, and the link will have to be engineered around these fixed points.

The site selection process for repeaters and receivers should also take into consideration the RF environment in which the equipment will be working. Multi-user locations, such as the World Trade Center in New York City or Mount Wilson near Los Angeles, are very good transmit sites, but terrible receive sites. For such situations, a remotely located receiver (in the case of a repeater system) should be considered. The two sections of the ARS station would then be tied together with telco facilities.

The use of a telco loop to feed a remotely located transmitter introduces several familiar problems, such as noise, crosstalk, distortion (if repeated several times), limited frequency response and installation and service delays. In some arrangements, the use of a telco loop to the transmitter (or receiver) is the only economical way to complete the link. Although such a hybrid system is not the ideal configuration, it will get the job done. Here again, the newer digital telco services can be helpful, because of their higher audio fidelity and freedom from crosstalk, and their ease in bidirectional, multi-line applications.

Planning for any RENG system should begin with an accurate, detailed U.S. Geological Survey (USGS) map covering the proposed path. Note should be made of any natural obstructions or Fresnel clearance obstructions (such as mountains, hills or vegetation) or manmade obstructions (such as buildings, water tanks or transmitting towers) in the proposed path. The transmitting and receiving antennas should be plotted so that a minimum of .6 Fresnel Zone clearance is obtained over 4/3 earth radius. Information on obtaining USGS maps may be found in Chapter 1.7, *Broadcast Standards and Recommended Practices*; Fresnel Zone clearance is covered in more detail in Chapter 6.2, *Microwave and STL Systems*.

When planning a RENG path, a profile drawing of the transmitting and receiving antenna site, the terrain and any obstructions in between, should be made on graph paper set to 4/3 earth radius. The use of such graph paper will compensate for the curvature of the earth and the normal refraction of VHF and UHF frequency signals when determining Fresnel Zone and obstruction clearance. Simple height above sea level is insufficient to determine whether a natural or manmade obstruction will interfere with the RENG signal on a long distance path. Once a proposed path has been drawn, a visual inspection must be made of the area for any problems that could degrade the performance of the system. Particular attention should be paid to items not documented on the USGS maps, such as buildings and towers.

The terrain from the transmitting antenna to the receiving antenna must be examined not only for obstructions but for reflection possibilities as well. A large body of water will usually cause problems for a RENG system operating in the UHF frequencies. If the water is an even number of Fresnel Zones from the direct path, signal attenuation will likely occur at the receiver. Temperature changes and tidal conditions will also have an effect. Likewise, thick vegetation or forested areas can be reflective to RF signals when wet, creating a similar (but not so troublesome) problem. Generally, the solution to reflection problems is to size of a briefcase and weighed 35 lbs. Today's gear is small, lightweight and can deliver excellent audio quality. Stations can now go into the field for news and special event programs and maintain studio-like sound. The audio quality of remote broadcasts is more important now than ever before because radio station transmission systems and consumer receivers are constantly improving. Further, the listening audience is becoming more discriminating and demanding of news operations. Today's competitive marketplace requires more than just sound from the field. It demands clean, quiet audio with good frequency response and low distortion. TV news has shown the public that this is possible and has conditioned consumers to expect it.

Radio channels used for RENG work are typically shared by multiple broadcasters within a market, and so receipt of a license is no guarantee of unlimited interference free operation. Indeed, an unused channel is the exception, not the rule, in most urban areas of the United States. The frequency coordination process is a complicated procedure that requires careful thought and planning, and generally a great deal of lead time. Broadcasters in a given geographical area rarely have to decide whether they wish to become involved in frequency coordination efforts. Usually the need for coordination is painfully obvious to all persons involved in RENG activity in the region.

The main driving force behind coordination efforts has been the Society of Broadcast Engineers (SBE), which has set up a National Frequency Coordinating Committee (NFCC) to encourage and support local coordination efforts, and to provide whatever support might be needed in this regard. The subject is discussed in detail in Chapter 1.5, *Frequency Coordination*. Spectrum congestion is a sad fact of life to many stations engaged in RENG today. Users must recognize that coordination is vital to the reliable operation of remote broadcast systems, since spectrum congestion will no doubt become worse in the future, rather than better.

Licensing Procedures

RENG work is done on two primary bands of frequencies set aside by the FCC for RPU operation. A number of frequency groups are allocated near 150 MHz and 450 MHz. Some assignments are also made on frequencies in the 25 MHz region. A particular broadcast station is not restricted to a maximum number of RPU systems that it may put into operation. The needs of the station and the budget available for equipment purchase are, instead, the major controlling factors.

Most RENG activity is currently centered in the 150 MHz and 450 MHz bands. In these slices of spectrum, three major license classifications exist: *Automatic Relay Station* (ARS), *Base Station* and *Remote Pickup Mobile Station*.

ARS systems are designed to receive program material on one frequency and retransmit on another. In this way, the average area of the RENG system can be extended considerably.

Base stations are, as might be expected, fixed position transmitters used for communication between the central point and one or more remote points. Base stations may, in the event of emergency conditions, be used as a program relay channel for Emergency Alert System information.

Remote Pickup Mobile Stations consist of vehicle-mounted and portable (hand carried) transmitters. They are usually licensed as a system in conjunction with a principal base station, or stations. Remote Pickup Mobile Station licenses generally specify a minimum and maximum number of mobile transmitters allowed in the RPU system. Standard divisions include from 1 to 4 stations, 4 to 12 stations, 10 to 20 stations and 20 to 50 stations.

The Commission's Rules require that the transmitter power for a RPU station be limited to a level necessary for satisfactory coverage of the service area. In any event, not more than 100 W of transmitter power output will be licensed. RPU transmitting equipment operating onboard an aircraft is normally limited to a maximum transmitter power of 15 W. A Mobile Station consisting of a hand carried or pack carried transmitter is restricted to not more than 2.5 W power output.

All RPU transmitting equipment must be type accepted by the Commission and checked each year (for units with more than 3 W output) for frequency accuracy, deviation and RF power output. FCC Rules also require that RPU transmitters rated for 3 W or greater must be equipped with a circuit that will automatically prevent modulation in excess of the authorized limits.

There are virtually no operator requirements for the use of a unit in the RPU service. Any person designated by and under the control of the licensee of the station may operate the equipment. An operator's license, as detailed in Part 13 of the Commission's Rules, is not required.

Building an RPU System

In view of the serious spectrum congestion problems that exist today in many areas of the country, any RPU/RENG system should be designed to be as spectrum efficient as possible and—equally important—to be as immune to undesired transmissions as possible. Even if the system will be operated in an area that currently does not have a spectrum congestion problem, there is no guarantee that such a problem will not surface in the near future. In any event, a well engineered system is also a spectrum efficient system.

The first rule of spectrum efficiency is to use only the effective radiated power (ERP) necessary to do the job. There is no justification for putting 15 W into the air when 5 W will provide the desired (or acceptable) signal-to-noise figure from the receiver. Ideally, all transmitters in a RENG system would therefore be equipped with continuously variable power output stages. The operator at the remote site would then run the transmitter with only enough power output to reach the required S/N figure at the receive (studio) point. With some types of units, this method of operation is

A program transmitted back to the studio via a standard POTS line—without any bandwidth extension—will usually be brief in duration, if for no other reason than the poor audio quality typical of such an arrangement. Spot news reports are common examples of this method of program return.

Small, battery powered mic-to-line amplifiers are available to drive dial-up telephones through direct connection to the tip and ring wires of the phone company cable or through clip leads at the handset microphone pins. The direct connection method of coupling is preferred over the handset connection because the former bypasses the telephone hybrid coil assembly with its associated level loss and possible distortion. But for this direct connection method, the device feeding the phone line—known as a *coupler*—should have the ability to seize the line (meaning to hold it open, in an off-hook condition), which the telephone instrument itself does when using the handset type connection. With such a *seizing coupler*, a call can be established with the telephone instrument, then the coupler can be switched on, and the handset of the telephone hung up, with the coupler holding the line open by itself. Many couplers (based on the original Western Electric *QKT* style) will *not* seize without another phone instrument off-hook on the same line. The advantage of the seizing coupler is twofold. First, the handset microphone will not contribute its output to the phone feed, so the signal transmitted will be only what comes from the amplifier or mixer plugged into the coupler's input. Second, the transmitted level of the audio signal will be higher without the instrument off-hook. This will maximize the signal-to-noise ratio of the feed.

Output level of the device feeding a phone line (whether by coupler or handset) should be carefully controlled, since the dynamic range of the dial-up network is somewhat limited. Transmit level should not exceed +8 dBm for program material, or 0 dBm for test tones. Be aware that the received end noise floor may be only 25–30 dB below this on local calls, and considerably less on some international long distance calls.

A program sent to the studio over dial-up telco lines using bandwidth extension equipment can provide impressive audio quality. Reasonably flat frequency response from 50 Hz to 5 kHz is possible in real-time using a 2-line system, or at half speed with a single-line system; 3-line systems can pass 7.5 kHz audio in real-time. The multi-band compandor systems used by the latter are required to reduce the dial-up network noise, which exhibits a rising characteristic with frequency and also becomes more subjectively noticeable as the high frequency cutoff of the system is raised.

These multi-line frequency shifting systems are falling into disfavor as POTS codecs become more popular. It is expected that frequency shifting on POTS lines will become obsolete as the cost effectiveness, audio quality and reliability of POTS codecs exceed the earlier technology.

Two- and 3-line frequency shifting systems or POTS codecs can also be operated at half-speed, with the expected effective doubling of their high frequency passband (one octave added), and a commensurate loss of an octave on the low end—usually a worthwhile tradeoff in terms of resulting audio quality. Thus the dial-up network can actually pass 15 kHz audio, with a 3-line frequency shifter or a POTS codec (on one line), at half-speed.

Note that none of these systems (with the possible exception of the single-line frequency extender) can operate reliably via cellular phone. This is due to a cellular phone line's bandwidth limitations and the potential change in level and frequency response that can occur at cell-boundary switches. Such cell switches can occur even when the remote site is a fixed location, due to moving obstructions in the path (such as a passing truck or bus), or atmospheric disturbances. (For more information on use of the POTS network, see Chapter 3.10, *Telephone Network Interfacing*.)

The other option for dial-up remotes is ISDN, which is becoming widely available. Many permanent remote sites (concert halls, convention centers, sports venues, etc.) are already wired with ISDN circuits ordered by broadcasters. In some areas, lead time for ordering new ISDN service is substantially longer than for POTS, and installation and service fees are also somewhat higher for ISDN. Nevertheless, ISDN is becoming the service of choice for on-demand, high-quality dial-up audio backhaul from frequently used remote sites. (For more information on use of the ISDN network, see Chapter 3.11, *Common Carrier Audio Program Services*.)

Setting Up Leased-line Remotes

The most critical element in using dedicated, leased line telco service for remote backhaul is getting the order placed properly and with sufficient advance notice. Today, T-1 service is the most often used of leased line service for broadcast backhaul, rapidly replacing the earlier analog *equalized broadcast loop* or *program circuit*. (ISDN service is also increasingly used in applications that previously required broadcast loops, particularly for longer backhauls and local events of short duration or infrequent recurrence.)

Changing tariffs continue to alter the comparative costs of these services (analog loops, ISDN and T-1). The comparison is further complicated by the inequality of service structure. For example, program loops are one-way paths while T-1 and ISDN are bidirectional; ISDN is pay-by-the-minute service while analog loops and T-1 are charged on a flat fee per month basis. Installation fees and space requirements for terminal equipment will also vary among these services.

For critical live remotes, installation of more than one service may be advisable for purposes of redundancy. (For more on digital services, see *Emerging Technologies* in this chapter, and Chapter 3.11, *Common Carrier Audio Program Services*.)

WIRELESS SYSTEMS

RENG has come a long way since the narrow-band walkie-talkie days, when a 5 W portable unit was the

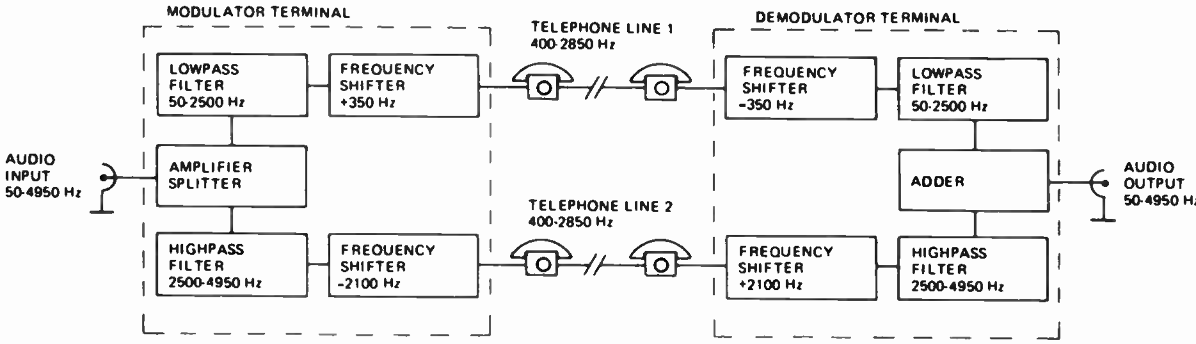


Figure 3.9-1. The basic 2-line frequency extension process for dial-up telephone links. (Courtesy of C. N. Rood.)

frequency-shifted back to their original values, filtered and recombined to form the output of the system. These multi-line systems may also include a multi-band compandor circuit for reducing phone line noise.

A useful variant of the single-line frequency extender employs *half-speed transmission*, as shown in Figure 3.9-2. Audio is recorded at the remote site on a two-speed analog tape recorder, using its higher recording speed. When the material is fed to the phone line, the deck's *lower* speed is used for playback and the audio is frequency-extension encoded by a single-line bandwidth extension system as it is played. At the receive end, the feed is decoded by the frequency extender and recorded using the low speed of another two-speed tape machine. When this program material is played back on high speed at the studio, an effective doubling of frequency response (a one octave increase) at the high end is achieved with a concurrent loss of one octave on the low end. The low end loss is usually of little consequence, since the frequency extension adds over two octaves, which is more than enough. This arrangement makes it possible to squeeze a signal bandwidth of 100 Hz-5 kHz through a single dial-up telco line. Phone feeding time takes twice as long, of course, and the process cannot be monitored for audio quality as the feed comes in. (For long feeds, it is therefore a good idea to make a short test recording and check its playback at the higher speed before proceeding with the entire program transmission.)

A more recent improvement to POTS audio transmission combines the technologies of computer modems with the perceptual coding (or *data compression* algorithms of the digital audio world. This class of devices, generically called *POTS codecs,* provides approximately 7 kHz audio transmission in real time on a single POTS line. The POTS codec turns a monaural analog audio input into a digital signal, applies a relatively heavy (around 30:1) data compression process—resulting in a datastream of about 24 kb/s—and feeds it to the POTS line via a standard 28.8 kb/s modem. At the receive end the process is reversed for a real-time audio output. (Actually, the term real-time in this context is inexact, since the throughput delay on a POTS codec path can be one second or more.) Some POTS codecs can adapt to variable phone line conditions and reduce their transmitted data rates to accommodate degraded line conditions—with corresponding reductions in audio bandwidth. In some cases (particularly for international calls), line quality is so poor that POTS codecs cannot operate reliably or at all. Some POTS codecs also include capacity to handle auxiliary control data or closures.

Setting Up a Dial-up Remote

If the dial-up (switched) option is chosen for a particular broadcast's backhaul, the broadcaster will have some, if not all, of the previously described service options available from the telco involved. The decision will depend upon the requirements of the particular broadcast and the costs of various services. Some generalizations can be made about these options, however, which will apply to most broadcast remote events.

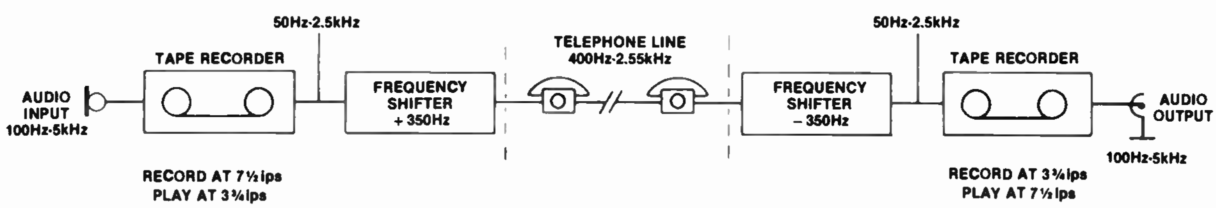


Figure 3.9-2. A frequency doubling technique using 2-speed tape recorders and a single-line bandwidth extension system. (Courtesy of C. N. Rood.)

company (telco) service. Spot news events, on the other hand, do not lend themselves to much advance planning. For these, an RF system is extremely useful.

The cost of telco facilities must also be considered. Unless the service is to be left in place for a long period of time, installation costs can become prohibitive, especially if high performance lines are needed. Many stations are able to justify the cost of a RENG RF system based solely on the telco savings that can be anticipated.

The best approach, therefore, implements both wired links and a wireless system. Large scale systems are often built using both interconnection methods, either as various links in the chain or as redundant services for recovery from a partial system failure.

While earlier dial-up (*switched*) and dedicated (leased line) telco services were exclusively analog, newer digital systems are becoming commonplace in most areas. For dedicated paths, the *T-1* service (also referred to as *DS1*) may be appropriate for broadcast use, particularly when multiple and/or bidirectional circuits are required between remote site and station. The T-1 format is a bidirectional digital service with a data transmission rate of 1.544 megabits per second (Mb/s) in each direction. It is primarily used by telcos to carry 24 voice-grade channels of 3.5 kHz bandwidth each, but can be easily reconfigured to carry a smaller number of wider-bandwidth signals by means of changing plug-in cards in terminal racks at each end of a dedicated T-1 path.

Program audio can be carried on T-1 in uncompressed or compressed digital form. A T-1 circuit can carry one uncompressed CD-quality stereo audio signal in each direction (occupying most of the path's 1.5 Mb/s bandwidth), while it may handle six or more such signals in compressed form. Alternatively, a mix-and-match configuration can be established with a variety of different bandwidths—a typical broadcast configuration between a studio and a permanent remote site (such as a news bureau or sports arena). For example, a T-1 link could be configured to carry several 15 kHz stereo program circuits bidirectionally (for transmission from, and monitoring returns back to, the remote site), bidirectional 5 kHz communication paths, bidirectional fax lines, voice circuits, and data control lines.

In some cases, a half-DS1 circuit or some other partial-carrier arrangement may be obtained, when capacity needs are not so great. Not available in all areas, this service is called *Fractional T-1*. In most cases, the T-1 digital service will be significantly more cost-effective than its analog program circuit equivalents (particularly when considering that the latter are not inherently bidirectional).

A more recent entrant to digital telco service is the *Integrated Switched Digital Network* (ISDN). ISDN is a dial-up digital service intended to replace the existing public switched telephone network. It uses telcos existing network of copper lines, and widespread, if not universal, ISDN deployment is expected by the end of the 1990s. The most common form of ISDN service, *basic rate interface* (BRI), provides two 64 kb/s channels for program audio. Using a technique called *inverse multiplexing* (IMUX), these two channels can be combined to provide a 128 kb/s path, which allows good stereo or excellent mono audio fidelity in a compressed digital form. ISDN therefore puts real time, high quality audio backhaul only a (digital) phone call away. Like POTS, either local or long-distance ISDN paths can be routed via dialing. ISDN fee schedules are also similar to or slightly higher than POTS service costs.

Note that broadcasters using T-1 or ISDN services for high quality audio transport must purchase or lease specialized terminal equipment (including codecs, when compressed audio is used), however. This can amount to several thousand dollars for each circuit, in addition to the one-time installation and ongoing monthly interconnection fees for digital service paid to telcos. (For more information on T-1 and ISDN, see Chapter 3.11, *Common Carrier Audio Program Services*.)

POTS Systems

For most remote backhaul needs, the conventional analog dial-up telephone network still gives today's broadcaster the greatest degree of flexibility in terms of when and where, including bidirectional ability. Of course, POTS audio quality leaves a great deal to be desired. For this reason, a number of encode-decode systems have been developed that improve the fidelity of audio signals that pass through a POTS line.

One such system that became quite popular during the 1980s is *frequency extension*. A variety of extension methods exist, each with a different way of accomplishing the task. Generally speaking, however, audio from the remote source is heterodyned up or down to frequencies that will pass through POTS circuits (which typically exhibit an approximate 300 Hz-3 kHz bandwidth), then detuned back to the original frequencies at the receive end. The one-decade response of the phone line can thereby be made to pass not *more*, but *a different set* of frequencies than it would normally carry.

The simplest of these systems shifts the effective passband of a phone line down by 250 Hz, making it approximately 50–2750 Hz. The process adds over two octaves to the low-end cutoff of the typical phone line, while only sacrificing a small portion (about 1/6) of an octave on the high end. This is a result of the heterodyne-shift's frequency-linear action, versus the logarithmic (octave-based) spectral perception of human hearing.

More sophisticated shifting systems add a second or even a third phone line to the process, using them to extend the *upper* frequency limit of the feed (see Figure 3.9-1). For example, in a 2-line unit, remote audio is split into two frequency bands by a filtering network. The lower frequency components are shifted upward by one conversion circuit (as in the basic system) and fed into one phone line, while the higher frequency components are shifted downward by a second conversion circuit, and applied to a second telephone line. At the receive end, the two signals are

3.9
REMOTE NEWS AND PRODUCTION

SKIP PIZZI
MICROSOFT CORPORATION, REDMOND, WA
JERRY WHITAKER
TECHNICAL PRESS, MORGAN HILL, CA

INTRODUCTION

Radio stations have used the remote location broadcast for decades to bring the listener an added sense of realism and excitement. Although the concept of the remote, as it is better known, has not changed substantially over the years, the means to accomplish the task has moved quantum leaps in terms of performance, ease of operation and reliability.

This chapter addresses the two most critical technical areas in remote broadcasting: 1) the portable production systems used to create the program (field production); and 2) the method of *backhaul* (the means of delivering the signal from the remote site to the broadcast studio location). The latter, often referred to as *radio electronic news gathering* (RENG), utilizes either telephone lines or RF links in the broadcast auxiliary or other spectrum.

Today's RENG systems can be configured to provide virtually any degree of sophistication required by the broadcaster. As with any other area of telecommunications, the key to a successful RENG system is thoughtful planning.

PLANNING THE RENG NETWORK

The importance of careful planning of a RENG system cannot be overemphasized. The network should be configured based on the precise needs of the broadcaster. Everyone involved in the use of the system should be consulted to determine what will be needed. Whether a broadcaster's format is *all news* or *all hits*, everyone in the news, production and engineering departments should sit down and define the requirements of the network. At such gatherings, engineers should resist the urge to be negative when someone asks for a level of performance that is not practical. Even though the engineers may know that it is impossible to provide every reporter on the staff with a separate frequency that can be received at the studio from anywhere in town, at least listen to what the users would like the system to do. The realities of station economics and the laws of physics can be explained after the desires of the participants have been outlined.

Many—perhaps most—RENG systems have been built on a piecemeal basis, as needs dictated and economics allowed. The lack of a unified plan has often led to RENG systems that are cumbersome to operate and, in the long run, more expensive to build than necessary for a given level of performance.

The size and layout of the broadcaster's market will have a substantial effect on how the RENG network is designed. A system intended to cover a sprawling urban area of 10,000 square miles will be configured much differently than a tightly clustered urban center covering 2,000 square miles. The number of broadcasters in the market that are involved in RENG activity may also affect how a system is designed, and what types of equipment are used. Broadcasters in major metropolitan areas may find that few, if any, frequencies are available for RENG activity.

Program material can be returned from the field to the studio through either of two common routes—wired telephone lines or wireless transmission systems of various types. The route back to the studio will depend upon a number of factors, including the location of the event, availability and type of telephone service in the area, amount of setup time provided and duration of the broadcast.

Wired vs. Wireless

Until the 1960s, the word *remote* was rarely spoken without reference in the same sentence to the telephone company. Wired systems, either using the dial-up *plain old telephone service* (POTS) network or leased broadcast loops, provided the vast majority of interconnections from remote broadcast sites to a station's studio facilities. Since that time, however, RF systems have assumed an important role in remote activities because they inherently offer greater flexibility and in many cases provide a higher quality audio link. Given sufficient use over time, they may also be more cost effective than leased program circuits from the telephone company.

Radio systems are ideally suited for broadcasts of relatively short duration and from several different locations during a short span of time. Longer lasting remotes, such as meetings, speeches and sporting events, are probably best handled by a wired arrangement. The amount of frequency congestion in the origination area will also have an effect on which method a station will choose for the greatest reliability. Urban areas in which secure *remote pickup unit* (RPU) channels are difficult to find are also good candidates for wired links.

The amount of lead time provided before a remote is scheduled to occur will also have a significant effect on the backhaul method chosen. Remotes scheduled weeks in advance are well suited for use of a telephone market stations, some degradation of the program (as perceived on a high-quality monitor) is often accepted for the sake of maximizing punch and loudness. Moderate quality compromises are usually masked on smaller and lower cost radios and are noticeable only on higher-quality radios by critical listeners.

In all cases: It is not appropriate to attempt to extrapolate the results of tone tests to program material, because superposition does not hold.

REFERENCES

1. British Broadcasting Corporation Engineering Division: "The dynamic characteristics of limiters for sound programme circuits," Research Report No. EL-5, 1967.
2. R.A. Haller, "An Update on the Technology of Loud Commercial Control," OST Technical Memorandum FCC/OST TM83-1, February 1983.
3. B.L. Jones and E. L. Torick, "A New Loudness Indicator for Use in Broadcasting," J. SMPTE, Sept. 1981, p. 772.
4. R. Orban, "Increasing Coverage of International Shortwave Broadcast Through Improved Audio Processing Techniques," *Journal of the Audio Engineering Society*, June 1990, p. 419.
5. SECAM customarily uses AM sound, with the usual requirements for preventing carrier pinch-off.
6. A phase-linear filter has constant delay with frequency.
7. A minimum phase filter has no zeros in the right half of the s-plane. As its name implies, there is no filter with the same magnitude response that can have less phase shift. Given the magnitude response of a minimum-phase filter, its phase shift can be computed (with the Hilbert Transform). This means that if a minimum-phase filter has constant group delay in its passband, this is associated with a certain type of magnitude response which rolls off gently around the filter's cutoff frequency: a minimum-phase filter with constant group delay in the passband cannot simultaneously have a highly selective magnitude response. Many textbooks provide the well-known mathematical details. See, for example, H.J. Blinchikoff & A.I. Zverev, *Filtering in the Time and Frequency Domains*, New York, Wiley, 1976, pp. 89-94.
8. Simply stated, almost all feedback systems contain a filter that forces the open loop characteristic to be either low-pass (all-pole, *lag compensation*) or low-pass shelving, with poles *and* zeros (*lead-lag compensation*). Feedback forces the amplifier before this filter to present a pre-emphasized signal to the filter's input such that the total response of the system is flat. If the filter rolls off at 6 dB/octave starting at 15 Hz (a typical situation in an opamp like the TL072 or the LF353), this pre-emphasis *rises* at 6 dB/octave 15 Hz. High frequencies applied to this system will obviously challenge the headroom of the amplifier prior to the filter. For example, 20 kHz will be up 62.4 dB! If high frequencies drive the amplifier prior to the filter into clipping or substantially nonlinear operation, transient intermodulation distortion occurs. Because a filter with a low-pass characteristic follows the clipping process, harmonics generated by clipping will be de-emphasized, so difference-frequency IM tests are more sensitive than THD tests to this mechanism.
9. For a maximum audio bandwidth f, the required slew rate in percent modulation per microsecond is $0.0002\pi f\%/\mu s$. For 4.5 kHz, this is 2.827%/μs.
10. West German (dual-carrier); Japanese (FM subcarrier); English NICAM (block-companded digital).

channels are left, center, right, left surround, right surround, and a limited-bandwidth subwoofer channel (the 0.1) for effects. The audio is digitally data-compressed using the Dolby AC-3 system. This system specifies two auxiliary data channels. The first, *dialog normalization*, provides information to the receiver about the nominal level of the dialog so that the receiver can hold this constant. This lets the broadcaster trade off headroom (for loud sound effects, for example) against noise floor without changing the loudness of dialog at the receiver.

The second auxiliary data channel is a dynamic range control channel that provides a wideband gain reduction signal that can be used by the receiver, under the viewer's control, to selectively compress the dynamic range of the broadcast. Unfortunately, at the time of this writing it is unclear how these signals will be provided operationally in a typical television broadcast environment, particularly with regard to older material mixed without these signals, and live news.

TECHNICAL EVALUATION OF AUDIO PROCESSING

Common swept frequency response, harmonic distortion, and intermodulation distortion tests are often used to evaluate audio processors. Therefore, it is useful to discuss why these tests may at times produce misleading results.

Definition of Linearity

A system can be tested for linearity as follows. Apply an input signal A to the system and measure its output. Let X be the output signal caused by input A. Then, remove A from the input and apply another signal B. Let Y be the output signal caused by the input B.

The system is linear if the following things happen: (1) If the input waveform is multiplied by a factor k to scale it, the output waveform also becomes scaled by a factor of k, but its shape is not distorted by the process of scaling. (2) If inputs A and B are applied to the system simultaneously, the system's output is X + Y (*superposition*).

It is clear that expanders, compressors and limiters are strongly nonlinear systems. The output of such a device is not scaled proportionally to its input; it is expanded or compressed. Similarly, when two signals are applied to such a device, its output is not the same as the sum of its response to either signal individually; superposition does not hold. Clippers are similarly nonlinear.

Sinewave Measurements and Non-Linearity

When predicting a system's response to program material by measuring its response to individual sinewaves, certain assumptions are made. The first assumption is that can be adequately represented program material as a sum of sinewaves (Fourier analysis). The second assumption is that superposition holds, so that the response of the system to single sinewaves also applies when several sinewaves are added together at the system's input. Thus, the sinewave results can be extrapolated to program material.

Because dynamic audio processing (compression, limiting, clipping, expansion, gating) is strongly nonlinear, the usual assumptions of superposition and scaling, which permit sinewave measurements to be extrapolated to complex program material through Fourier analysis, do not hold. Conventional harmonic and intermodulation distortion measurements, historically designed to measure slight departures from linearity in weakly nonlinear systems, are of very limited usefulness. Swept or spot frequency response measurements are not useful.

When making distortion measurements with tones, their relevance must be assessed psychoacoustically. Does the system output *sound* distorted when listening to the tones? For example, when measuring harmonic distortion using fundamentals in the 50–1000 Hz region, the higher harmonics are more significant than the lower harmonics because the higher harmonics are less readily masked by the desired fundamental. However, as the fundamental frequency is increased, the harmonics become less troublesome because the ear becomes less and less sensitive to them. Eventually, their frequency exceeds the passband of the system and they become irrelevant.

Similarly, SMPTE intermodulation distortion methods measures the level of 50 or 60Hz sidebands around a high-frequency tone induced by system nonlinearity. Because these sidebands are within a single critical band (approximately 1/3 octave) of the high-frequency tone, they are maximally masked by it. Therefore rather high amounts of measured SMPTE IM distortion are not necessarily cause for concern. On the other hand, ITU-R difference-frequency intermodulation distortion measurements measure the low-frequency difference tone caused by two high-frequency tones. Because the difference tone is far removed in frequency from the desired tones, it is not well masked by them, and high amounts of ITU-R IM are of some concern. See Chapter 7.1, "Audio Signal Analysis," for more information on distortion and measurements.

Subjective Listening Tests

There are few, if any, measurement techniques that can adequately predict whether the subjective effect of an audio processor will be satisfactory. The only effective way to evaluate nonlinear broadcast audio processing is by *subjective listening tests*. These must be done over a long time period, using many different types of program material, because a processor that sounds good on a certain type of program material may sound unsatisfactory on other program material having markedly dissimilar spectral balance or dynamics.

Usually, the subjective goal of broadcast processing is to have its action undetectable to the audience. In the case of processing in highly competitive major

AUDIO PROCESSING REQUIREMENTS FOR DIGITAL AUDIO RADIO (DAR) TRANSMISSION

At this writing, the question of which system the United States will adopt for DAR broadcasting is unsettled. Most of the rest of the world has already standardized on the Eureka 147 system. To encode the audio, all systems use lossy data reduction with no pre-emphasis. The specific system of lossy data reduction that is used depends on the DAR system, with Eureka 147 using the MPEG-1 Layer 2 algorithm, and some other proposed systems using the Lucent/AT&T PAC algorithm or the MPEG AAC algorithm.

The Eureka 147 system provides an auxiliary data channel that can be used as a gain reduction signal to provide dynamic range compression at the receiver under the listener's control. Because this provides wideband compression, it cannot provide a processed sound without overt spectral gain intermodulation. Indeed, it is intended only to provide gentle AGC to make the signal comfortably listenable when the radio is turned down to background levels or when it is located in an acoustically noisy environment.

As of this writing, there is no history of commercial engineering practice in DAR broadcasting, as the pioneer broadcasters have principally been government and public radio services. Nevertheless, based on the experience of FM, the author expects that some stations will want to apply substantial multiband audio processing at the transmitter. Such processing will perform no high frequency limiting (because there is no pre-emphasis). Peak limiting can be similar to the technology used for compact disc mastering, which often uses delay-line limiters.

One important consideration in processing for DAR is the possibility of the lossy data reduction decoder's clipping at the receiver because of overshoots added by the data reduction. An aggressively peak-limited signal will spend considerable time close to digital full-scale, and, due to the addition of quantization noise, the lossy algorithms will introduce considerable overshoots above this level. At this writing, most of the decoder algorithms have no headroom built-in to accommodate such overshoots, instead clipping them. The author hopes that this oversight will be corrected before DAR receivers hit the mass market.

AUDIO PROCESSING REQUIREMENTS FOR TELEVISION BROADCAST TRANSMISSION

The processor should provide a comfortably listenable dynamic range in domestic listening environments by applying subtle compression to the signal. Such compression should be undetectable to the ear unless the original source is available for comparison. Usually an available gain reduction range of 25 dB is adequate to handle the level variations encountered in typical operations.

The processor must provide high frequency limiting to complement the pre-emphasis employed (50 μsec or 75 μsec, depending upon the region in which the transmission occurs). The processor should provide accurate peak control (as measured by a modulation monitor meeting the standards of the governing authority) in both the positive and negative directions. In general, the comments on FM (given previously) apply here as well.

The processor should control subjective loudness to prevent unpleasant inconsistencies when transitions occur between various program elements. This is most accurately achieved using technology similar to that developed for loudness measurement (discussed earlier). In essence, the processor uses a loudness meter in a servo loop to control loudness and ensure consistency of loudness between one program source and the next.

The processor should handle voice cleanly. The Hilbert-Transform clipper and delay-line limiter are effective for this, because neither creates audible clipping distortion on voice, even when the source is narrowband (such as optical film or telephone). Such narrowband sources are extremely difficult for a conventional audio-frequency clipper to process without introducing some audible harmonic distortion on voice.

Audio Processing for Stereo Television

The general requirements for stereo television processing are not very different from the general requirements enumerated previously. As discussed in the *Processing for Stereo* section earlier in this chapter, the processing elements with slow release time constants must be coupled to preserve the stability of the stereo image. In the North American BTSC system, the peak modulation criteria are complex. However, it can be shown that FM stereo-style processing will always prevent over-modulation in BTSC stereo, although it will not necessarily allow the most L + R modulation theoretically possible in this system. This style of processing is also appropriate for the other international stereo systems,[10] since it will always prevent over-modulation.

Because of the close proximity between the edge of the audio passband (approximately 15 kHz) and the stereo pilot tone (15.734 kHz), the BTSC system requires sharp low-pass filters to prevent aliasing. It is impractical at the current state of the art to apply nonlinear overshoot compensation to these filters. Such overshoots do not cause interference or problems in television receivers. Thus these overshoots must be accepted as inherent to the BTSC system, and must be ignored by modulation monitors designed as a reference for setting modulation levels. If these overshoots are not ignored, average modulation will be set too low and the viewer will experience annoying increases in loudness when switching from stereophonic to monophonic channels.

The new advanced digital television system specified by the Grand Alliance and approved by the FCC for use in the U.S. specifies 5.1 channel audio. The

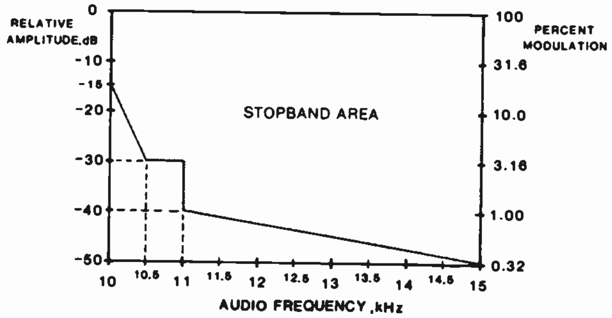


Figure 3.8-5. NRS low-pass filter curve.

audio bandwidth to 5 kHz at night. This will cause little or no audible change in frequency balance as heard on the average AM radio (whose audio bandwidth at −3 dB continues to be approximately 3 kHz, despite the push by broadcasters for AMAX radios), but it will completely prevent any interference from being applied to first-adjacent channels. (At night, skywave propagation can convey such interference to stations hundreds of miles from the interferer.) In addition, the station is likely to find that it can process its audio harder for a given amount of processor-induced distortion, increasing on-air loudness and coverage. This is particularly important at night because interference increases (see Figure 3.8-5).

On April 27, 1989, The FCC released a *Report and Order* that amended section 73.44 of the FCC Rules by requiring all U.S. AM stations to comply with the occupied bandwidth specifications of the NRSC-2 standard by June 30, 1990. The NRSC-2 standard is an *RF mask* which was derived from the NRSC-1 audio standard by the NRSC. The purpose of the NRSC-2 RF mask is to provide a transmitted RF occupied bandwidth standard that any station with a properly-operating transmitter will meet, *provided that NRSC-1 audio processing is used prior to the transmitter*, and *provided that the station is not overmodulating.*

AUDIO PROCESSING FOR AM STEREO

In all AM stereo systems, the envelope modulation is forced to a close approximation of the sum of the left and right channels to ensure compatibility with mono radios equipped with envelope detectors. To ensure minimum loudness loss compared to monophonic transmission, it is necessary to process stereo audio in the *sum and difference* format. This means that the left and right channels are passed through matrix circuits to create L + R (sum) and L − R (difference) signals. These signals are then passed separately through those parts of the processing that control modulation.

To prevent clipping and distortion in the C-Quam decoder in the receiver, the negative-going modulation in the left and right channels must be no greater than −75% modulation (where 100% modulation is carrier pinch-off). Therefore, an audio processor for C-Quam must have both a sum and difference processor (which does the main processing) and a safety limiter to protect the left and right channels. This safety limiter is usually inactive, and typically only comes into play when the input program has sections which are momentarily single-channel, such as ping-pong stereo.

AUDIO PROCESSING REQUIREMENTS FOR FM (VHF) BROADCAST TRANSMISSION

The processor should provide a comfortably listenable dynamic range in domestic and automotive listening environments by applying subtle compression to the signal. Unless the program director requests otherwise for competitive reasons, such compression should be undetectable to the ear unless the original source is available for comparison.

The processor must provide high frequency limiting to complement the pre-emphasis employed (50 μsec or 75 μsec, depending upon the region in which the transmission occurs).

The processor must provide accurate peak control (as measured by a modulation monitor meeting the standards of the governing authority) in both the positive and negative directions. To ensure that absolute peak control will be retained at the system output, any system elements following the processor must have flat frequency response (±0.1 dB) and constant group delay (deviation from linear phase <±10°). Because the pre-emphasis networks and low-pass filters ordinarily found in stereo encoders do not meet these requirements, they should be bypassed. Thus, the processor should provide pre-emphasis and band limiting for the transmission system. Its output must contain negligible energy above the bandwidth limit of the transmission system. In FM stereo broadcasting by the world-standard pilot-tone method, this bandwidth is limited to less than 19 kHz to prevent aliasing from the stereo subchannel into the main channel, and vice-versa. To protect the pilot tone itself (ensuring correct operation of the phase-locked loop subcarrier regeneration circuitry in the receiver's stereo decoder), the bandwidth must be further limited to no greater than 17 kHz. In practice, it is customary to begin the HF rolloff at slightly above 15 kHz to minimize group delay distortion in the low-pass filters used to effect the bandwidth limit. Nonlinear low-pass filters are usually used to prevent overshoot, enabling the processor to control peak deviation absolutely.

The processing system must be readily adjustable to achieve the subjective effect desired by the broadcasting authority operating it. To achieve a competitive sound in markets where many stations compete for listeners, it may be necessary to add additional multiband limiting to the basic audio processing system (which usually consists of compressor, HF limiter, and peak limiter/clipper). Adding additional multiband limiting can create greater program density than the basic processing system alone without introducing spectral gain intermodulation.

amplification stage prior to the frequency compensation stage) probable.[8] To minimize the probability that TIM will be bothersome, any amplification stage before the frequency compensation stage should be designed to be very linear to its clipping point, and to have sufficiently high headroom to accommodate the maximum rate of change to be expected at the transmitter's audio input.[9]

A transmitter can be qualified for TIM by one of the various difference-frequency intermodulation distortion tests. If the tests indicate that the transmitter has a low slew rate, it will not respond well to pre-emphasized audio and pre-emphasis will have to be reduced until the first derivative of the processed audio waveform seldom, if ever, exceeds the slew rate limit of the transmitter. Because of the benefits of pre-emphasis at the receiver, it is desirable to modify such transmitters to increase their slew rate, even if this means somewhat compromising harmonic distortion performance at low frequencies.

The NRSC-1 Audio Standard

As the North American AM band became more crowded, interference from first and second adjacent stations became more of a problem. Receiver manufacturers responded by producing receivers with decreased audio bandwidth, so that the encroachment of an adjacent station's modulation extremes would not be audible as interference.

This truncating of the bandwidth had the effect of diminishing the receiver's high-frequency response, but it was felt that lower fidelity would be less annoying than interference. To address these problems, the National Radio Systems Committee (NRSC) in 1987 formalized a standard for pre-emphasis and low-pass filtering for AM broadcast to provide brighter sound at the receiver while minimizing interference. See Chapter 4.3, "AM Transmitters," for more information on NRSC work.

AM Stereo Introduces a Pre-emphasis Dilemma

Certain AM receivers manufactured since 1984 for sale in North America, particularly those designed for domestic AM stereo reception, have a frequency response that is substantially wider than that of the typical mono AM receiver. The frequency response was widened largely to enhance the sales potential of AM stereo by presenting a dramatic, audible improvement in fidelity in the showroom. As these new receivers became more prevalent, broadcasters had to choose whether the station's pre-emphasis would be optimized for the new AM stereo receivers or for the existing conventional receivers that form the vast majority of the market.

If the choice was for conventional receivers (which implies a relatively extreme pre-emphasis), the newer receivers might sound strident or exceptionally bright. If the choice favored the newer receivers (less pre-emphasis and probably less processing), the majority of receivers would be deprived of much high-end energy and would sound *duller* and have less loudness.

NRSC Standard Pre-emphasis and Low-pass Filtering

In response to this dilemma, the NRSC undertook the difficult task of defining a voluntary recommended pre-emphasis curve for AM radio that would be acceptable to broadcasters (who want the highest quality sound on the majority of their listeners' radios) and to receiver manufacturers (who are primarily concerned with interference from first- and second-adjacent stations).

A modified 75-microsecond pre-emphasis/de-emphasis standard was approved (see Figure 3.8-4). That provides a moderate amount of improvement for existing narrowband radios, while optimizing the sound of wideband radios. Most importantly, it generates substantially less first-adjacent interference than do steeper pre-emphasis curves.

The second part of the NRSC standard calls for a sharp upper limit of 10 kHz for the audio presented to the transmitter (see Figure 3.8-5). This essentially eliminates interference to second and higher adjacencies. While some broadcasters believe that this is inadequate and that 15 kHz audio should be permitted, it is not likely that interference-free 15 kHz audio could be achieved except by a reallocation of the AM band. The practical effect of widespread implementation of the 10 kHz standard is that 10 kHz radios are feasible, and the bandwidth perceived by the average consumer (now typically limited by the receiver to 3 kHz) can be dramatically improved. The difference between AM and FM reception will then become less pronounced.

The NRSC and the EIA worked together to define a standard for wideband, NRSC-compliant AM stereo radios, which was trademarked as AMAX. Several manufacturers (most notably Delco) have introduced AMAX radios into the marketplace.

Although the NRSC standard calls for a 10 kHz audio bandwidth at all times, one can seriously argue that any North American AM station whose programming is primarily talk should voluntarily reduce its

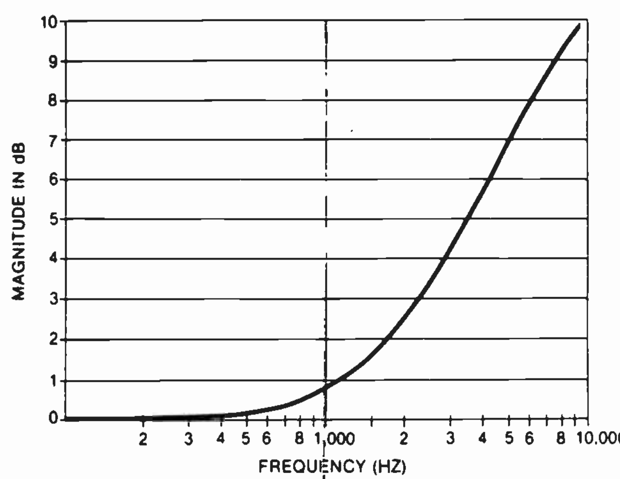


Figure 3.8-4. NRSC pre-emphasis curve.

cessor must compensate for noise (electrical and acoustic) and interference by reducing dynamic range. This is most readily done by multiband compression and limiting to achieve lowest peak-to-average ratio without significant processing-induced side effects.

The processor must provide absolute negative peak control to prevent AM carrier pinch-off, which would otherwise cause out-of-band emissions. Additionally, the processor must incorporate overshoot-free filtering to control the audio input spectrum to the transmitter, thus preventing out-of-band emissions and interference. National (FCC) or international broadcast authorities (most notably ITU-R) usually specify the permissible occupied bandwidth to make most efficient use of available radio frequency spectrum.

The processor may also be equipped with a receiver equalizer that compensates for the poor frequency response of the typical MW or SW radio due to narrowband RF and intermediate frequency (IF) stages.

Transmitter Equalization

The processor may provide a transmitter equalizer to eliminate tilt, overshoot, and ringing in the transmitter and antenna. Accurate reproduction of the shape of the processed waveform requires that the transfer function between the audio input and the modulated RF envelope represent a constant delay (which may be any positive number or 0) at all frequencies contained within the audio input signal. Failure to meet this criterion can result in tilt, overshoot, and *ringing* in the modulated RF envelope. The cause of overshoot and ringing as spectrum truncation and time dispersion at the high-frequency end of the system bandpass were discussed earlier. Tilt, on the other hand, is caused by problems at low frequencies.

Figure 3.8-3 shows the response of a 10 kW plate-modulated transmitter to a 50 Hz square wave. The transmitter causes the waveform to tilt, which increases peak modulation in both positive and negative directions. The magnitude of the transmitter's frequency response is essentially flat to 50 Hz; the problem is caused by infrasonic rolloff. This rolloff is equivalent to that of a high-pass filter and is minimum-phase, which introduces time dispersion, causing the shape of the waveform to change and further increasing the peak level.

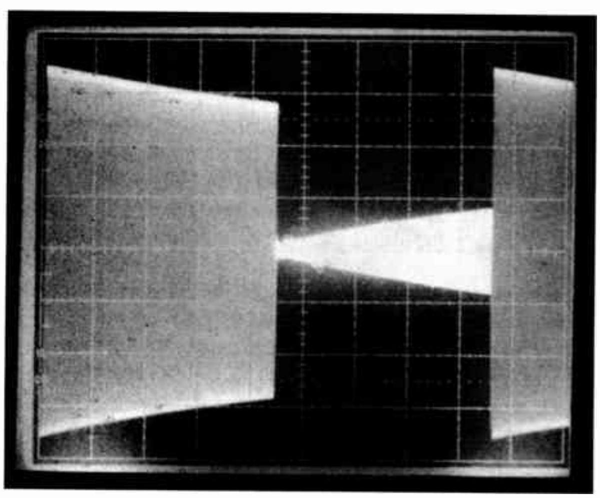


Figure 3.8-3. Tilt in plate modulated transmitter.

Some transmitters contain high-pass filters at their audio inputs to protect high-power stages. *This location is absolutely inappropriate;* these filters can easily increase the peak-to-average ratio of the input audio by 3-4 dB. The correct location for a protection high-pass filter is in the audio processor where measures can be taken to prevent the high-pass filter from increasing the peak-to-average ratio at the audio processor's output.

Bounce

Predistorting the waveform in the audio processor can equalize linear errors. However, one major nonlinear error, commonly called *power supply bounce* is caused by resonances in the LC filter elements of the transmitter's high voltage power supply. These resonances superimpose a sub-audible modulation onto the power supply voltage, resulting in a form of very fast carrier shift that is too quick to be seen on a conventional carrier shift meter. The net result is to compromise the control of modulation peaks, particularly on strong bass transients which cause momentarily large current demands on the power supply, and which excite the resonance.

In some older transmitters, bounce has been known to compromise achievable modulation by up to 3 dB. Because bounce is not linearly related to the modulation, small-signal equalization cannot cure it. The most successful cure has been the use of a 12-phase power supply in the transmitter. The AC ripple from such a supply is down about 40 dB without filtering; a simple filter capacitor is all that is necessary to achieve adequate smoothing. Because there are no chokes in the power supply filter, resonance cannot occur. In all cases, bounce can be minimized by preventing excessive bass energy from being applied to the transmitter.

Slew Rate Limiting (Transient Intermodulation Distortion)

Transmitters using pulse-duration modulation (PDM) schemes are prone to problems with slew rate limiting. Because the PDM low-pass filter is located within the audio feedback loop of the transmitter, and because this filter is typically a multi-pole elliptic function filter with a cutoff frequency below 70 kHz, it will introduce substantial delay into the feedback loop. This has two consequences: stability requires the amount of feedback applied around the transmitter to be limited, and it also requires that the open-loop gain of the modulator be rolled-off at a very low frequency. The first issue makes it difficult to design PDM transmitters with THD below 1%–2% at midrange frequencies, while the second renders *transient intermodulation distortion* (TIM non-linear behavior of the can cause noise that would otherwise be masked to become unmasked because the psychoacoustic masking conditions under which the masking thresholds were originally computed have changed. Second, the frequency response of the radio receiver (particularly in AM) can remove frequencies that were used to make the psychoacoustic masking calculations and that would otherwise have masked the added quantization noise.

Accordingly, if you use lossy data reduction in the studio you should use the highest data rate possible. This maximizes the headroom between the added noise and the threshold where it will be heard. Also, you should minimize the number of encode/decode cycles, because each cycle moves the added noise closer to the threshold where it will be heard.

Transmission Levels and Metering

Engineers at the transmitter and the studio consider transmission levels and their measurements differently. Transmission engineers need to know the peak level of a transmission commonly measured by an oscilloscope. Studio engineers need to know the line-up (or reference) level of a transmission commonly measured by a VU meter (as the approximate RMS level) or by a peak program meter (as the PPM level). For details, see the standard established by the Institute of Electrical and Electronic Engineers, *Recommended Practice for Audio Program Level Measurement* (Doc G-2.1.2/13, 1988).

Metering

The VU meter is an average-responding meter (measuring the approximate RMS level) with a 300 ms rise time and decay time; the VU indication usually lags the true peak level by 8 to 14 dB. PPM indicates a level between RMS and the actual peak. The PPM reading has an attack time of 10 ms, slow enough to cause the meter to ignore narrow peaks and lag the true peak level by 5 dB or more.

Transmission Levels

The transmission engineer is primarily concerned with the peak overload level of a transmission to prevent overloading. This peak overload level is defined differently, system to system. In tape, it is defined as the level producing the amount of harmonic distortion considered *tolerable*—often 3% THD at 400 Hz. In FM, microwave, or satellite links, it is the maximum-permitted RF carrier deviation. In AM, it is negative carrier pinch-off. In analog telephone transmission, it is the level above which serious crosstalk into other channels occurs, or the level at which the amplifiers in the channel overload. In digital, it is the largest possible digital word.

Studio Levels

The studio engineer is primarily concerned with what is commonly called the reference level, operating level, or line-up level. This line-up level aids studio engineers in providing adequate headroom between line-up level and the overload level of equipment to allow for the peaks that the meter does not indicate. In facilities that use VU meters, line-up level is usually at 0 VU, which corresponds to the studio standard level, typically +4 or +8 dBm. In systems that use PPM, line-up level may be at PPM 4 (for the BBC standard) or at the studio standard level (often +6 dBm).

Transmission-Link Limiting

Transmission-link limiting devices are sometimes used ahead of the transmission link to protect it from overload. (These links might be STLs, satellite uplinks, inter-studio digital links etc.) These devices are usually used below-threshold (that is, with no gain reduction) as protection limiters to control peak levels. They only produce gain reduction when abnormally high levels are applied to their input due to operator error or unforeseen level variations at the source. This is useful to transmission engineers concerned with overload, and as useful to studio engineers concerned with headroom. For the needs of both engineers, such a limiter's output must be adjusted to be at or slightly below the peak overload level of the transmission channel.

To properly match the studio line-up level to the transmission protection limiter, the desired headroom must be known. For example, assume that the transmission protection limiter produces 0 dBm at its output at 100% modulation of the transmission link. Further, assume that the line-up level in a production facility is designed to allow 8 dB of headroom. The input attenuator of the transmission protection limiter would then be adjusted so that studio line-up tone produces −8 dBm at the output of the transmission protection limiter.

This assumes that the amplifier or other link between the studio and the input of the transmission protection limiter has enough headroom to drive the transmission protection limiter into gain reduction without clipping this link. The transmission protection limiter only protects a link connected to its output. In the previous example, if the transmission protection limiter provides 15 dB of maximum protection, the system prior to the transmission protection limiter requires 8 + 15 = 23 dB of headroom above studio line-up level. If the link is simply an amplifier, this should be achievable without difficulty if the absolute level of the studio line-up tone is chosen carefully. In our example, if the amplifiers in the system clip at +21 dBm, the level of the studio line-up tone can be no greater than −2 dBm (i.e., 23 dB below +21 dBm).

AUDIO PROCESSING REQUIREMENTS FOR MW AND HF BROADCAST TRANSMISSION

In amplitude-modulated services, reception is usually compromised by noise and interference and may be further compromised by acoustic noise in the listening environment (such as the automobile). Thus the pro-

Where only an audio link is available, feed the audio output of the processing system directly into the link. If possible, transmitter protection limiters should be adjusted for minimum possible action as the processing system does most of that work. Transmitter protection limiters should respond only to signals caused by faults or by spurious peaks introduced by imperfections in the link.

Where maximum quality is desired, it is important that all equipment in the signal path after the studio be carefully aligned and qualified to meet the appropriate standards for bandwidth, distortion, group delay, and gain stability, and that such equipment is requalified at reasonable intervals.

Requirements for STLs

If the STL is prior to the audio processor, the STL's signal-to-noise ratio (SNR) must be sufficient to pass unprocessed audio. This means that the SNR of the link must be better than the sum of the desired SNR of the transmitted signal plus the maximum gain of the audio processor plus about 6 dB (a useful rule-of-thumb). If the STL follows the audio processor, its SNR must be 6 dB better than the desired SNR of the transmitted signal. To ensure that the STL does not distort the shape of the audio waveform (preventing introduction of overshoot into peak-limited waveforms applied to the STL input), the frequency response must be flat (±0.1 dB) throughout the operating frequency range, typically 20–15,000 Hz. The group delay must be essentially constant throughout this range (deviation from linear phase <±10°). Phase correction can be applied to meet the requirement at high frequencies.

At low frequencies, by far the best way to achieve the specification is to extend the −3 dB frequency of the STL to 0.15 Hz or lower and to eliminate any peaking in the infrasonic frequency response prior to the rolloff. It is not unusual for a microwave STL to bounce because of a large infrasonic peak in its frequency response caused by an under-damped automatic frequency control (AFC) phase-locked loop. This bounce can increase the STL's peak carrier deviation by as much as 2 dB, reducing average modulation. Many commercial STLs have this problem. Such problems can be corrected by modifying the STL's AFC loop, or by applying equalization prior to the STL transmitter that is complementary to existing low-frequency rolloff, such that the overall system frequency response rolls off smoothly at 0.15 Hz or below. This solution is far better than clipping the tilt-induced overshoots after the STL receiver because the clipping will introduce nonlinear distortion, while the equalizer is distortion-free.

For highest quality, the nonlinear distortion of the STL system should be less than 0.1% total harmonic distortion (THD) throughout the operating frequency range.

Digital links may pass audio as straightforward Pulse Code Modulation (PCM) encoding, or they may apply data-rate-reduction processing to the signal to reduce the number of bps required for transmission through the digital link. Such processing will almost invariably distort peak levels, and such links must therefore be carefully qualified before you use them to carry the peak-controlled output of the audio processor to the transmitter or stereo encoder. For example, the MPEG-1 Layer 2 algorithm can increase peak levels up to 4 dB at 160 kB/sec by adding large amounts of quantization noise to the signal. While the desired program material may psychoacoustically mask this noise, it is nevertheless large enough to affect peak levels severely. For any lossy compression system, the higher the data rate, the less the peak levels will be corrupted by added noise, so use the highest data rate practical in the system. Even with the maximum available data rate, overshoot will probably be large enough to require use of an overshoot compensator at the STL receiver (or in the equipment being driven by it). Some modern FM exciters and stand-alone stereo encoders are now equipped with such compensators.

Other links may use straightforward PCM without lossy data rate reduction. These can be very transparent and can exhibit accurate pulse response provided that their input anti-aliasing filters and output reconstruction filters are rigorously designed to achieve constant group delay over the frequency range that contains significant program energy. This is not particularly difficult to do with modern over-sampled converter technology.

NICAM is essentially a hybrid between PCM and data-rate-reduced systems. It uses a block-companded floating-point representation of the signal with J.17 pre-emphasis. NICAM links can exhibit low overshoot if designed for good low-frequency response and equipped with phase-linear anti-aliasing and reconstruction filters.

Because the output spectrum of most modern audio processing systems is already tightly band-limited, you may bypass any anti-aliasing filters in digital links driven by such systems. This ensures the most accurate possible transient response.

Using Lossy Data Reduction in the Studio

Many stations are now using lossy data reduction systems to increase storage time of digital playback media. In addition, source material is often supplied through a lossy data reduction system, whether from satellite or over landlines. Sometimes, several encode/decode cycles will be cascaded before the material is finally presented to the audio processor's input.

All such algorithms operate by increasing the quantization noise in discrete frequency bands. If not psychoacoustically masked by the program material, this noise may be perceived as distortion, *gurgling*, or other interference. Psychoacoustic calculations are used to ensure that the added noise is masked by the program material and cannot be heard. In addition, there are at least two other mechanisms in broadcasting that can cause the noise to become audible at the radio. First, a modern multiband transmission audio processor performs an *automatic equalization* function that can radically change the frequency balance of the program. This frequency bands have slower time constants than the high-frequency bands. This multiband compressor usually does most of the work in increasing program density.

The amount of *gain reduction* determines how much the loudness of soft passages will be increased (and, therefore, how consistent overall loudness will be). Our hypothetical system reduces gain with the broadband AGC and the multiband compressor. The broadband AGC is designed to control average levels, and to compensate for a reasonable amount of operator error. It is *not* designed to substantially increase the short-term program density; the multiband compressor and peak limiters handle that function.

Modern audio processing systems usually add other elements to the basic system described here. For example, it is not unusual to incorporate an equalizer to color the audio for artistic effect. The equalizer may be any of the types described earlier and is usually found between the slow AGC and the multiband compressor. The multiband compressor itself can also be used as an equalizer by adjusting the gains of its various bands.

Various low-pass filters are often included in the system to limit the bandwidth of the output signal to 15 kHz (for FM), 10 kHz (AM in NRSC countries), 4.5 kHz (AM in EBU countries, and shortwave worldwide) or other bandwidths as required by the local regulatory authority. The final low-pass filter in the system is almost always overshoot-compensated to prevent introducing spurious modulation peaks into the output waveform. High-pass filters may be incorporated to protect the transmitter. This is particularly important in high-power AM and shortwave installations exceeding 100 kW carrier power.

A transmitter equalizer that corrects the pulse response of the transmitter is found on high-end AM processors.

Location of System Components

The best location for the processing system is as close as possible to the transmitter, so that the processing system's output can be connected to the transmitter through a circuit path that introduces the least possible change in the shape of the carefully peak-limited waveform at the processing system's output. Sometimes, it is impractical to locate the processing system at the transmitter, and it must instead be located on the studio side of the link connecting the audio plant to the transmitter. (The studio/transmitter link (STL) might be telephone or post lines, analog microwave radio, or various types of digital paths.) This situation is not ideal because artifacts that cannot be controlled by the audio processor can be introduced in the link to the transmitter or by additional peak limiters placed at the transmitter. (Such additional peak limiters are common in countries where the transmitter is operated by a different authority than that providing the broadcast program.)

In this case, the audio output of the processing system should be fed directly to the transmitter through a link that is as flat and phase-linear as possible. Deviation from flatness and phase-linearity will cause spurious modulation peaks because the shape of the peak-limited waveform is changed. Such peaks add nothing to average modulation. Thus the average modulation must be lowered to accommodate those peaks within the carrier deviation limits dictated by government authorities.

This implies that if the transmitter has built-in high-pass or low-pass filters (as some do), these filters *must* be bypassed to achieve accurate waveform fidelity. Modern processing systems contain filters that are fully able to protect the transmitter, but which are located in the processing system where they do not degrade control of peak modulation.

Where Access To The Transmitter Is Available

The audio received at the transmitter site should be of as good quality as possible. Because the audio processor controls peaks, it is not important that the audio link (STL) feeding the processing system's input terminals be phase-linear. However, the link should have low noise, flattest possible frequency response from 30–15,000 Hz, and low nonlinear distortion.

If the audio link between the studio and the transmitter is noisy, performing the compression function at the studio site can minimize the audibility of this noise. Compression applied before the audio link improves the signal-to-noise ratio because the average level on the link will be greater. If the STL has limited dynamic range, it may be desirable to compress the signal at the *studio* end of the STL. To apply such compression, split the processing system, placing the AGC and multiband compressor sections at the studio, and the peak limiter at the transmitter.

Where Access To The Transmitter Plant Is Not Available

In some situations, the organization originating the program does not have access to the transmitter, which is operated by a separate entity. In this case, all audio processing must be done at the studio, and any damage that occurs later must be tolerated. A peak limiter would, however, be used at the transmitter to provide protection against overmodulation.

If it is possible to obtain a broadband phase-linear link to the transmitter, the processing system at the studio location can feed the STL. The output of the STL receiver is then fed directly into the transmitter with no intervening processing. A *composite* STL (ordinarily used for FM stereo baseband) has the requisite characteristics, and can be used to carry the output of the processing system to the transmitter. However, the output of a typical composite STL receiver is at the wrong level and impedance to directly drive a typical transmitter (most of which require +10 dBm into 600 Ω). Therefore, the transmitter must almost certainly be modified to make it compatible with the composite STL. Use of a composite STL has many ramifications, and the installation of the processing system at the transmitter may be less complicated.

changing—a characteristic side-effect of wideband compressors and limiters when driven heavily), *breathing* (audible pulling up of background noise, cured by a compressor gate), and *hole-punching* (a sudden drop in loudness after a program transient, caused by the transient's inducing a large amount of gain reduction which then does not decay quickly, and cured by multiple time-constant release time circuitry).

The processor must be packaged so that it is easy to operate and maintain, and can work in high RF fields without compromise.

The processor should have setup controls with enough versatility to enable the subjective effect to be readily tuned to the requirements of the broadcasting authority operating it. For mixed-format applications the processor may have several presets, selectable by remote control, that permit the operator to set the amount of compression, limiting, clipping and other parameters to complement the program material being transmitted.

Ordinarily, the processor should be equipped with sufficient remote control facilities to enable it to be interfaced efficiently with modern, automated plants. Most of the required facilities are specific to the application: for example, medium wave (MW), shortwave (HF), FM (VHF), or television.

The processor should have sufficient metering to permit it to be easily set up with tones or program material. The metering should also provide operations and diagnostic capabilities. Metering usually includes input level, output level, and gain reduction (the amount of limiting or compression) occurring in each variable-gain stage.

Processing for Stereo

Processing for stereophonic transmission is similar to processing for monophonic transmission, except that two audio processing chains are used. To preserve stereo imaging, the gains of the left and right automatic gain control and compression circuitry must be identical. Conversely, experience has shown that fast peak limiting and high-frequency limiting circuits sound best when operated independently (without stereo coupling), because the ear does not perceive channel-imbalance-induced spatial shifts with these fast time constants. However, the ear can perceive the loudness of one channel's being modulated unnaturally by a dominant element in the other channel when the channels are coupled.

The gain of the coupled elements is determined by the requirements of the transmission service. In FM, the channel requiring the greatest amount of limiting determines the gain of both channels. The processor operates by sensing the higher of the left and right channels and determining the gain of *both* channels such that the higher channel does not exceed a given level at the processor's output.

In AM, the gain of both channels is controlled by sensing and controlling the level of their sum (L + R), because the envelope modulation represents the sum of the channels.

SYSTEM CONSIDERATIONS

Reducing the peak-to-average ratio of the audio increases *loudness*. If peaks are reduced, the average level can be increased within the permitted modulation limits. The level with which this can be accomplished without introducing objectionable side effects (like clipping distortion) is the single best measure of audio processing effectiveness. *Density* is the extent to which the amplitudes of audio signal peaks are made uniform (at the expense of dynamic range). Programs with large amounts of short-term dynamic range have low density and highly compressed programs have high density.

Compression reduces the difference in level between the soft and loud sounds to make more efficient use of permitted peak level limits, resulting in a subjective increase in the loudness of soft sounds. It *cannot* make loud sounds seem louder. Compression reduces dynamic range relatively slowly in a manner similar to "riding the gain." Limiting and clipping, on the other hand, reduce the short-term peak-to-average ratio of the audio.

Limiting increases audio density. Increasing density can make loud sounds seem louder, but can also result in an unattractive, busier, flatter and denser sound. It is important to be aware of the many negative subjective side effects of excessive density when setting controls that affect the density of the processed sound.

Clipping sharp peaks does not produce any audible side effects when done moderately. Excessive clipping will be perceived as audible distortion.

Building a System

Combining several audio processors into a good sounding system is tricky because of headroom and time constant considerations. The device driving a given processor must be able to drive that processor into full compression or limiting. If the driving device (for example, distribution amplifier) runs out of headroom before full limiting occurs in the driven device, then that device cannot achieve its full capability. This consideration is particularly critical when setting up the input A/D converter of a digital audio processor. The analog drive level to the A/D converter must be set so that the A/D converter does not clip when receiving levels sufficient to cause full gain reduction in the following audio processor.

Beware of interactions between the attack times and release times when cascading several processors. It is wise to start the system with the slowest device. This is usually a compressor or automatic gain controller (AGC) with slow attack and release times and a compressor gate to prevent noise breathing. Such a processor does not significantly increase the *density* of the audio; it simply does gentle gain riding to ensure that following stages are driven at the correct level.

A multiband compressor with moderate attack and release times often follows the slow AGC. Correctly designed multiband processors have these time constants optimized for each frequency band. The low-well to perceived loudness. A meter that purports to measure loudness must agree with a panel of human listeners.

There are three important factors that correlate to subjective loudness:

- The spectral distribution of the sound energy. The ear's sensitivity depends strongly on frequency. It is most sensitive to frequencies between 2 and 8 kHz. Sensitivity falls off fastest below 200 Hz.
- Whether the sound energy is concentrated in a wide or narrow bandwidth. For a given total sound power, the sound becomes louder as the power is spread over a larger number of *critical bands* (about 1/3 octave). This is called *loudness summation*.
- The duration of the sound. A given amount of sound power appears progressively louder until its duration exceeds about 200 nsec, at which point no further loudness increase will occur.

Torick and Jones have published a paper describing a meter for measuring the loudness of broadcast signals.[3] The FCC did an informal validation of the results of this meter, and concluded that it was effective in assessing whether commercials in television were noticeably louder than the surrounding entertainment programming.[4]

Additionally, the independently developed loudness measuring methods of Stevens and of Zwicker have both become international standards. At the time of this writing there is at least one commercially available loudness meter using the Jones & Torick method. This can be used as an approximate measure of subjective loudness. Bear in mind that members of Jones & Torick's listening panel disagreed on the loudness of certain sounds by over 3 dB, so any metering represents an average and many not agree exactly with loudness as perceived by any one individual.

GENERAL PERFORMANCE REQUIREMENTS FOR TRANSMISSION AUDIO PROCESSORS

The audio processor must control the peak modulation of the RF carrier to the standards required by the governing authority, such as the FCC in the United States. In AM, this usually means that negative carrier pinch-off must not occur at any time because this would cause splatter interference into adjacent channels. In FM and television (NTSC and PAL), the peak deviation of the carrier must be controlled so that the modulation monitor specified by the governing authority does not indicate overmodulation.[5] Because the rules often permit the modulation monitor to ignore very brief overshoots, the instantaneous peak deviation might exceed the peak modulation as indicated on the modulation monitor.

The requirements for peak control and spectrum control tend to conflict, which is why sophisticated nonlinear filters are required to achieve highest performance. Applying a peak-controlled signal to a linear filter almost always causes the filter to overshoot and ring because of two mechanisms: *spectral truncation* and *time dispersion*. One can build a square wave by summing its Fourier components together with correct amplitude and phase. Analysis shows that the fundamental of the square wave is approximately 2.1 dB higher than the amplitude of the square wave itself. As each harmonic is added in turn to the fundamental, a given harmonic's phase is such that the peak amplitude of the resulting waveform *decreases* by the largest possible amount. Simultaneously, the RMS value *increases* because of the addition of the power in each harmonic. This is the fundamental theoretical reason why simple clipping is such a powerful tool for improving the peak-to-average ratio of broadcast audio: clipping adds to the audio waveform spectral components whose phase and amplitude are precisely correct to minimize the waveform's peak level while simultaneously increasing the power in the waveform.

If a square wave (or clipped waveform) is applied to a low-pass filter with constant time delay at all frequencies, the higher harmonics that reduce the peak level will be removed, increasing the peak level and with it the peak-to-average ratio. Thus even a perfectly phase-linear low-pass filter will cause overshoot.[6] *There is no sharp-cutoff linear low-pass filter that is overshoot-free:* overshoot-free spectral control to FCC or ITU-R standards must be achieved with filters that are embedded within the processing, such that the nonlinear peak controlling elements in the processor can also control the overshoot.

If the sharp-cutoff filter is now allowed to be minimum-phase, it will exhibit a sharp peak in group delay around its cutoff frequency.[7] Because the filter is no longer phase-linear, it will not only remove the higher harmonics required to minimize peak levels, but will also change the time relationship between the lower harmonics and the fundamental. They become delayed by different amounts of time, causing the shape of the waveform to change. This *time dispersion* will therefore further increase the peak level.

When a square wave is applied to a linear-phase filter, overshoot and ringing will appear symmetrically on the leading and trailing edge of the waveform. If the filter is minimum phase, the overshoot will appear on the trailing edge and will be about twice as large. In the first case, the overshoot and ringing are in fact caused by spectrum truncation which eliminates harmonics necessary to minimize the peak level of the wave at all times; in the second case, the overshoot and ringing are caused by spectrum truncation *and* by distortion of the time relationship between the remaining Fourier components in the wave.

Application Considerations

Except as required to achieve very specific artistic goals (most notably in some major-market high-energy hit-music formats), the processed audio should be free from unnatural subjective side-effects, such as *pumping* (a sense that the gain is constantly and unnaturally

The *noise gate* is an expander with a high expansion ratio. Its purpose is to reduce noise. Because it causes gain reduction when the input level drops below a given threshold, the ear is likely to hear the accompanying gain reduction as a fluctuation in the noise level, sometimes called *breathing*. This can sound unnatural. Therefore, the noise gate is most useful when applied to a single microphone in a multi-microphone recording. Usually, the other microphones will mask any breathing, yet the noise reduction provided by the noise gate will still be appreciated during quiet program material.

Multiband Compression and Frequency-Selective Limiting

These techniques divide the audio spectrum into several frequency bands and compress or limit each band separately (although some interband coupling may be used to prevent excessive disparity between the gains of adjacent bands). This is the most powerful and popular contemporary audio processing technique, because, when done correctly, it eliminates spectral gain intermodulation. This occurs in a wideband compressor or limiter when a voice or instrument in one frequency range dominates the spectral energy, thus determining the amount of gain reduction. If other, weaker, elements are also present, their loudness may be audibly and disturbingly modulated by the dominant element. Particularly unpleasant effects may occur if the dominant energy is in the bass region, because the ear is relatively insensitive to bass energy, so the loudness of the midrange is pushed down by the dominant bass energy seemingly inexplicably.

Another type of frequency-selective limiting uses a program-controlled filter. The filter's cutoff frequency, its depth of shelving, or a combination of these parameters, is varied to dynamically change the frequency response of the transmission channel. Such program-controlled filters are most often used as *high-frequency limiters* to control potential overload due to pre-emphasis in pre-emphasized systems like FM (VHF), and television audio (NTSC and PAL), and in FM modulated transmission channels such as microwave links and satellite circuits.

Equalization

Equalization is changing the spectral balance of an audio signal, and is achieved by use of an *equalizer*. In broadest terms, an equalizer is any frequency-selective network (filter) placed in the signal path. In audio processing, an equalizer is usually a device that can apply a *shelving* or *peaking* curve to the audio.

A shelving curve starts off at a certain gain. As frequency changes, the gain increases (boost) or decreases (cut) asymptotically. Finally, the gain shelves off and does not change with further changes in frequency.

A peaking curve is bell-shaped on the frequency axis. As opposed to a shelving curve, it has a well-defined peak frequency. The shape of the curve can be uniquely defined by three *parameters: the amount of equalization* (in dB), *the frequency of maximum equalization* (in Hz) and the *Q*, which is a dimensionless number that describes whether the curve is broad or sharp.

A *parametric equalizer* provides several peaking equalizers, in which the user has control of all three parameters. This type of equalizer is generally considered to be the most flexible and musical-sounding equalizer. Some parametric equalizers can also be used as notch filters.

A *graphic equalizer* provides a number of peaking equalizers (usually 8 to 31) distributed on octave or fractions of octave (1/4 or 1/3) spaced frequency centers throughout the audible range. The controls for the amount of equalization are linear-throw faders, and are arranged on the panel in order of frequency. The positions of the controls, when considered together, thus provide a very rough graphic display of the amount of equalization provided by the entire equalizer. The advantage of a graphic equalizer is that it is easy to understand and quick to adjust. Its primary disadvantage is lack of flexibility. Usually, only the amount of equalization is adjustable, the Q and center frequency being fixed. However, a few manufacturers make parametric equalizers with graphic-style controls. These provide the advantages of both types.

Low-pass and high-pass filters remove spectrum at the top and bottom of the audible range, respectively. They are usually used to remove unwanted high- or low frequency noise, and can also produce special effects (like telephone simulation).

These filters come with their rate of cutoff fixed in multiples of 6 dB/octave. 12 dB/octave and 18 dB/octave are popular. In addition, the shape of the region around the cutoff frequency has a considerable effect on the listening quality of such filters. Bessel (constant-delay) filters have a gentle transition into cutoff, and sound pleasant and musically neutral. Butterworth (maximally-flat magnitude) filters have a sharper transition into cutoff. They are more effective at removing noise than Bessel filters, but have a more colored listening quality.

Equalizers are sometimes used on-line in transmission to create a certain sonic signature for a broadcast. Any of the types above may be used. Commercial audio processors may include equalizers for program coloration, or for correcting the frequency response of previous or subsequent transmission links. Sometimes the various bands of a multiband compressor or limiter are used as an equalizer by adjusting the gains of the various bands to achieve the desired equalized frequency response.

Loudness

One of the main uses of audio processing is to increase perceived loudness within the peak modulation constraints of a transmission channel. Assessing the effectiveness of audio processing thus requires a means of measuring loudness. *Loudness is subjective: it is the intensity of sound as perceived by the ear/brain system*. No simple meter, whether peak program meter (PPM) or VU, provides a reading that correlates

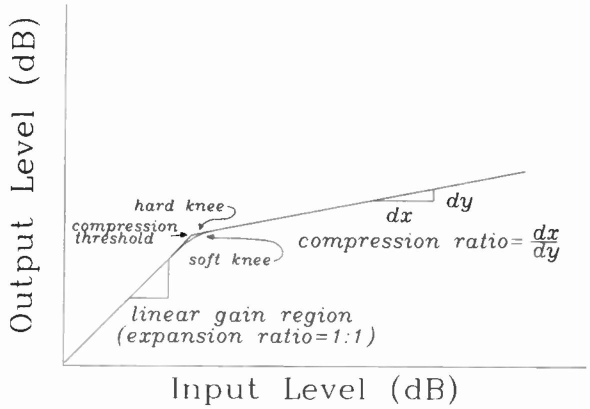


Figure 3.8-1 Input versus output levels for compressors.

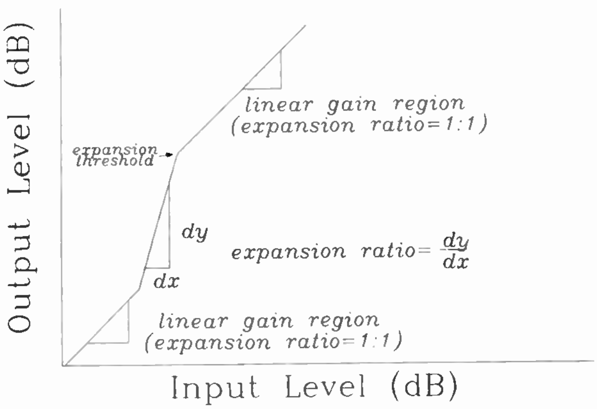


Figure 3.8-2. Input versus output levels for expanders.

attack time (typically less than 2 μsec), and fast release time (typically less than 200 μsec). In modern audio processing, a peak limiter, by itself, usually limits the peaks of the *envelope* of the waveform, as opposed to individual instantaneous peaks *in the waveform*. These are usually controlled by *clipping*. As a matter of good engineering practice, peak limiters are usually adjusted to produce no more than 6 dB of gain reduction to prevent offensive audible side effects.

The main purpose of limiting is to protect a subsequent channel from overload, as opposed to compression, the main purpose of which is to reduce the dynamic range of the program.

Peak clipping is a process that instantaneously chops off any part of the waveform that exceeds the *threshold of clipping*. This threshold can be either symmetrical or asymmetrical around 0 volts. While peak clipping can be very effective, it causes audible distortion when over used. It also increases the bandwidth of the signal by introducing both harmonic and intermodulation distortion into its output signal. Manufacturers of modern audio processors have therefore developed various forms of *overshoot compensation*, which is essentially peak clipping that does not introduce significant out-of-band spectral energy into its output.

Radio frequency clipping (RF clipping) is peak clipping applied to a single-sideband RF signal. (A typical carrier frequency is 1 MHz). All clipping-induced harmonics fall around harmonics of the carrier (2 MHz, . . .). Upon demodulation, these harmonics remain at high frequencies and are removed by a low-pass filter. Thus RF clipping produces only intermodulation distortion, and no harmonic distortion. Ordinary or audio-frequency (AF) clipping produces both. RF clipping is substantially more effective than AF clipping on voice because intermodulation distortion is considered less objectionable than harmonic distortion in this application. On the other hand, RF clipping is considered much more objectionable than AF clipping on music.

The *Hilbert Transform clipper* combines the features of RF and AF clippers.[1] It acts as an RF clipper below 4 kHz (the region in which most voice energy is located), and acts as an AF clipper above 4 kHz to prevent excessive intermodulation distortion with music.

Unless a limiter has an attack time of less than about 10 μsec, it will exhibit *overshoots* at its output. If the goal of the processing is to precisely constrain the instantaneous values of the waveform to a given threshold, it usually sounds best to control these overshoots by a limiter with 2 nsec attack time followed by a clipper. Attempting to provide all peak control with the limiter does not sound as good, because the clipper affects only the offending overshoot and does not apply gain reduction to the surrounding signal.

When used in this way, clippers can cause audible distortion on certain program material. However, fast-attack limiters will cause audible clipping of the first half-cycle of certain program material, such as solo piano, harp, and nylon string acoustic guitar. Such distortion can be eliminated by a *delay line limiter*.[2] This device consists of two audio paths. The audio is applied to a pilot limiter, which has a very fast attack time. The gain control voltage generated by the pilot limiter is applied to a low-pass filter to smooth its sharp edges, and then to the gain control port of the variable gain amplifier that passes the actual program signal. To compensate for the group delay in the control voltage low-pass filter (which delays application of gain control to the audio), the audio is delayed equally by a delay line prior to the variable gain amplifier's input.

Gating

There are two fundamental types of gates, the *compressor gate* and the *noise gate*. The *compressor gate* prevents any change in background noise during pauses or low-level program material by freezing the compressor gain when the input level drops below the *threshold of gating*. Because it produces natural sound, it is very popular in broadcasting.

Instead of freezing, many compressor gates will cause the gain to move very slowly to a nominal value (typically 10 dB of gain reduction) if the gating period is long enough. This prevents the compressor from getting stuck with an unusually high or low gain.

3.8
TRANSMISSION AUDIO PROCESSING

ROBERT ORBAN
ORBAN, INC., A HARMAN INTERNATIONAL COMPANY, SAN LEANDRO, CA

INTRODUCTION

Transmission audio processing is both an engineering and artistic discipline. The engineering goal is to make most efficient use of the signal-to-noise ratio and audio bandwidth available from the transmission channel while preventing its overmodulation. The artistic goal is set by the organization using audio processing. It may be to avoid audibly modifying the original program material at all. Or it may be to create a distinct *sonic signature* for the broadcast by radically changing the sound of the original. Most broadcasters operate somewhere in between these two extremes.

Provided that the transmitted signal meets regulatory requirements for modulation control and RF bandwidth, there is no well defined right or wrong way to process audio. Like most areas requiring subjective, artistic judgement, processing is highly controversial and likely to provoke exceedingly opinionated arguments between its practitioners. Ultimately, the success of a broadcast's audio processing must be judged by its results—if the broadcast gets the desired audience, then the processing must be deemed satisfactory regardless of the opinions of audiophiles, purists, or others who consider processing an unnecessary evil.

One mark of the professionalism of a broadcast engineer is their mastery of the techniques of audio processing. The canny practitioner has a bag of tricks that can be used to achieve the processing goal specified by the station's management, whether it is purist or "squashed against the wall."

FUNDAMENTALS OF AUDIO PROCESSING

Compression reduces the dynamic range of program material by reducing the gain of material whose average or root mean square (RMS) level exceeds the *threshold of compression.* The amount by which the gain is reduced is called the *gain reduction* (G/R).

Above threshold, the slope of the input/output curve is the *compression ratio.* Low ratios provide loose control over levels, but generally sound more natural than high ratios, which provide tight control.

The *knee* of the input/output level graph can show an abrupt transition (*hard-knee*) into compression, or a gradual transition (*soft-knee*), in which the ratio becomes progressively larger as the amount of gain reduction increases.

The *attack time* is, generally, the time that it takes the compressor to settle to a new gain following a step increase in level. There is no generally agreed upon precise definition on how to measure attack time. Some measure it as the *time constant*—the time necessary for the gain to achieve 67% of its new value. Others measure it as the time for the gain to reach 90% of its new value for a given amplitude step (often 10 dB).

The *release time* is the time necessary for the gain to recover to within a certain percent of its final value after the level of the input signal to the compressor has been reduced below the compression threshold. It is sometimes convenient to specify the release time in dB per second if the shape of the release time is a straight line on a dB versus time graph. However, this shape often is not linear. *Multiple time constant* (sometimes called automatic) release time circuits change the release rate (in dB/second) according to the history of the program, and according to how much gain reduction is in use. For example, the release time will temporarily speed up after an abrupt transient, to prevent a hole from being punched in the program by the gain reduction. The release time may slow down as 0 dB gain reduction is approached to make compression of wide-dynamic-range program material less obvious to the ear.

Delayed release holds the gain constant for a short time (typically less than 20) after gain reduction has occurred. This prevents fast release times from causing modulation of individual cycles in the program waveform, thus reducing the tendency of the compressor to introduce harmonic or intermodulation distortion when operated with fast attack and release. The foregoing compression parameters are illustrated in Figure 3.8-1.

Expansion

Expansion increases the dynamic range of program material by reducing gain when the program level is lower than the *threshold of expansion* (see Figure 3.8-2). The primary purpose of expansion is to reduce noise, either electronic or acoustic. Expanders are often coupled to compressors so that low-level program material is not amplified, thus reducing the noise that would otherwise be exaggerated by the compression. Expanders have attack times, release times and expansion ratios that are analogous to those for compressors.

Peak Limiting and Clipping

Peak limiting is an extreme form of compression characterized by a very high compression ratio, fast

Todd, C., et. al., "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," *AES 96th Convention, Preprint 3796,* Audio Engineering Society, New York, February 1994.

Wylie, Fred., "Audio Compression Techniques," *The Electronics Handbook,* Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 1260–1272, 1996.

Zwicker, E., "Subdivision of the Audible Frequency Range Into Critical Bands (Frequenzgruppen)," *J. Acoust. Soc. of Am.*, vol. 33, p. 248, February 1961.

to assessing the overall performance of an audio broadcasting system.

The problem with subjective evaluations is that, while accurate, they are time consuming and expensive to undertake. Traditional objective benchmarks of audio performance, such as signal-to-noise ratio or total harmonic distortion, are not reliable measures of perceived audio quality, especially when perceptually coded signals are being considered.

To remedy this situation, ITU-R established Task Group 10-4 to develop a method of objectively assessing perceived audio quality. Conceptually, the result of this effort would be a device having two inputs—a reference and the audio signal to be evaluated—and would generate an audio quality estimate based on these sources.

Six organizations proposed models for accomplishing this objective, and over the course of several years these models were evaluated for effectiveness, in part by using source material from previously documented subjective evaluations. Ultimately, the task group decided that none of the models by themselves fully met the stated requirements. The group decided, instead, to use the best parts of the different models to create another model that would meet the sought after requirements.

This approach resulted in an objective measurement method known as *perceptual evaluation of audio quality* (PEAO). The method contains two versions—a basic version designed to support real-time implementations, and an advanced version optimized for the highest accuracy but not necessarily implementable in real-time. The primary applications for PEAQ are summarized in Table 3.7-2.

PERSPECTIVE ON AUDIO COMPRESSION

A balance must be struck between the degree of compression available and the level of distortion that can be tolerated, whether the result of a single coding pass or the result of a number of passes, as would be experienced in a complex audio chain or network. There have been many outstanding successes for digital audio data compression in communications and storage, and as long as the limitations of the various compression systems are fully understood, successful implementations will continue to grow in number.

Compression is a trade off and in the end you get what you pay for. Quality must be measured against the coding algorithm being used, the compression ratio, bit rate and coding delay resulting from the process.

There is continued progress in expanding the arithmetical capabilities of digital signal processors, and the supporting hardware developments would seem to be following a parallel course. It is possible to obtain a single chip containing both encoder and decoder elements, including stereo capabilities. In every five year period, it is not unreasonable to expect a ten fold increase in the processing capabilities of a single DSP chip, thus, increasing flexibility and processing power. Speculation could point to an eventual position when a completely lossless algorithm with an extremely high compression ratio would become available. In any event, the art of compressing audio data streams into narrower and narrower digital pipes will undoubtedly continue.

Table 3.7-2
Target applications for ITU-R Rec. BS.1116 PEAQ.

Category	Application	Version
Diagnostic	Assessment of implementations	Both
	Equipment or connection status	Advanced
	Codec Identification	Both
Operational	Perceptual quality line-up	Basic
	On-line monitoring	Basic
Development	Codec development	Both
	Network planning	Both
	Aid to subjective assessment	Advanced

REFERENCES

1. ATSC, *Digital Audio Compression Standard (AC-3),* Advanced Television Systems Committee, Washington, D.C., Doc. A/52, Dec. 20, 1995.
2. ATSC, *Digital Television Standard,* Advanced Television Systems Committee, Washington, D.C., Doc. A/53, Sep.16, 1995.

BIBLIOGRAPHY

Bonomi, Mauro, "The Art and Science of Digital Video Compression," *NAB Broadcast Engineering Conference Proceedings,* National Association of Broadcasters, Washington, D.C., pp. 7–14, 1995.

Brandenburg, K., and Gerhard Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *92nd AES Convention Proceedings,* Audio Engineering Society, New York, N.Y., 1992, revised 1994.

Ehmer, R. H., "Masking Patterns of Tones," *J. Acoust. Soc. Am.,* vol. 31, pp. 1115–1120, August 1959.

Ehmer, R. H., "Masking of Tones Vs. Noise Bands," *J. Acoust. Soc. Am.,* vol. 31, pp. 1253–1256, September 1959.

ITU-R Recommendation BS-775, *Multi-Channel Stereophonic Sound System With and Without Accompanying Picture.*

Moore, B. C. J., and B. R. Glasberg, "Formulae Describing Frequency Selectivity as a Function of Frequency and Level, and Their Use in Calculating Excitation Patterns," *Hearing Research,* vol. 28, pp. 209–225, 1987.

Smyth, Stephen, "Digital Audio Data Compression," *Broadcast Engineering,* Intertec Publishing, Overland Park, Kan., February 1992.

Stallings, William, *ISDN and Broadband ISDN,* 2nd Ed., MacMillan, New York.

Taylor, P., "Broadcast Quality and Compression," *Broadcast Engineering,* Intertec Publishing, Overland Park, Kan., p. 46, October 1995.

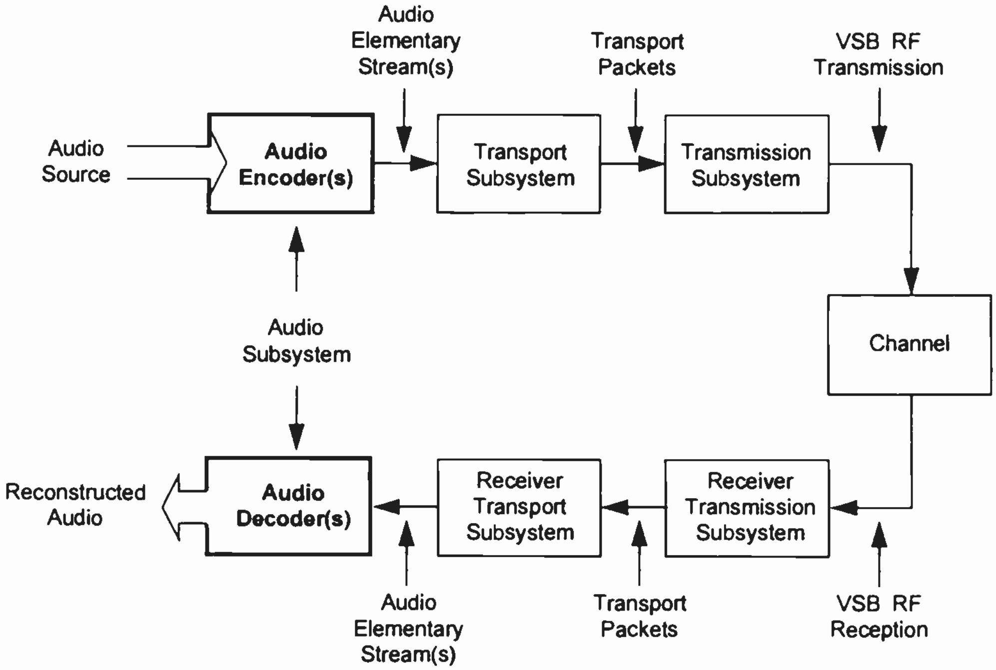


Figure 3.7-11. The audio subsystem in the DTV standard.[2]

An audio program source is encoded by a *digital television audio encoder*. The output of the audio encoder is a string of bits that represent the audio source (the audio elementary stream). The transport subsystem packetizes the audio data into *packetized elementary system* (PES) packets, which are then further packetized into transport packets. The transmission subsystem converts the transport packets into a modulated RF signal for transmission to the receiver. At the receiver, the signal is demodulated by the receiver transmission subsystem. The receiver transport subsystem converts the received audio packets back into an audio elementary stream, which is decoded by the digital television audio decoder.

The partitioning shown in Figure 3.7-11 is conceptual and practical implementations may differ. For example, the transport processing may be broken into two blocks; the first would perform PES packetization and the second would perform transport packetization. Or, some of the transport functionality may be included in either the audio coder or the transmission subsystem.

Sampling Parameters

The AC-3 system conveys digital audio sampled at a frequency of 48 kHz, locked to the 27 MHz system clock.[2] If analog signal inputs are employed, the A/D converters should sample at 48 kHz. If digital inputs are employed, the input sampling rate should be 48 kHz, or the audio encoder should contain sampling rate converters that translate the sampling rate to 48 kHz. The sampling rate at the input to the audio encoder must be locked to the video clock for proper operation of the audio subsystem.

In general, input signals should be quantized to at least 16 bit resolution. The audio compression system can convey audio signals with up to 24 bit resolution.

Objective Quality Measurements

Perceptual audio coding has revolutionized the processing and distribution of digital audio signals. One aspect of this technology, not often emphasized, is the difficulty of determining, objectively, the quality of perceptually coded signals. Audio professionals could greatly benefit from an objective approach to signal characterization because it would offer a simple but accurate approach for verification of good audio quality within a given facility.

Most of the discussions regarding this topic involve reference to the results of subjective evaluations of audio quality, where for example, groups of listeners compare reference audio material to coded audio material and then judge the level of impairment caused by the coding process. A procedure for this process has been standardized in ITU-R Rec. BS.1116, and makes use of the ITU-R five grade impairment scale:

- 5.0—Imperceptible
- 4.0—Perceptible but not annoying
- 3.0—Slightly annoying
- 2.0—Annoying
- 1.0—Very annoying

Quality measurements made with properly executed subjective evaluations are widely accepted and have been used for a variety of purposes, from determining which of a group of perceptual coders performs best, by the core bit allocation routine, which determines how many bits should be used to encode each individual mantissa. The spectral envelope and the coarsely quantized mantissas for six audio blocks (1536 audio samples) are formatted into an AC-3 frame. The AC-3 bit stream is a sequence of AC-3 frames.

The actual AC-3 encoder is more complex than shown in the simplified system of Figure 3.7-9. The following functions also are included:

- A frame header is attached, containing information (bit rate, sample rate, number of encoded channels and other data) required to synchronize to and decode the encoded bit stream
- Error-detection codes are inserted to allow the decoder to verify that a received frame of data is error-free
- The analysis filterbank spectral resolution may be dynamically altered to better match the time/frequency characteristic of each audio block
- The spectral envelope may be encoded with variable time/frequency resolution
- A more complex bit allocation may be performed, and parameters of the core bit allocation routine may be modified to produce a more optimum bit allocation
- The channels may be coupled at high frequencies to achieve higher coding gain for operation at lower bit rates
- In the 2-channel mode, a rematrixing process may be selectively performed to provide additional coding gain, and to allow improved results to be obtained in the event that the 2-channel signal is decoded with a matrix surround decoder.

Decoding

The decoding process is, essentially, the inverse of the encoding process.[1] The basic decoder, shown in Figure 3.7-10, must synchronize to the encoded bit stream, check for errors and deformat the various types of data (the encoded spectral envelope and the quantized mantissas). The bit allocation routine is run, and the results are used to unpack and dequantize the mantissas. The spectral envelope is decoded to produce the exponents. The exponents and mantissas are transformed back into the time domain to produce the decoded PCM time samples. Additional steps in the audio decoding process include the following:

- Error concealment or muting may be applied in the event a data error is detected
- Channels that have had their high frequency content coupled must be decoupled
- Dematrixing must be applied (in the 2-channel mode) whenever the channels have been rematrixed
- The synthesis filterbank resolution must be dynamically altered in the same manner as the encoder analysis filterbank was altered during the encoding process

Implementation of the AC-3 System

As illustrated in Figure 3.7-11, the audio subsystem of the ATSC DTV Standard comprises the audio encoding/decoding function and resides between the audio inputs/outputs and the transport subsystem.[2] The audio encoder is responsible for generating the *audio elementary stream*, which is an encoded representation of the baseband audio input signals. (Note that more than one audio encoder may be used in a system.) The flexibility of the transport system allows multiple audio elementary streams to be delivered to the receiver. At the receiver, the transport subsystem is responsible for selecting which audio streams to deliver to the audio subsystem. The audio subsystem is then responsible for decoding the audio elementary stream back into baseband audio.

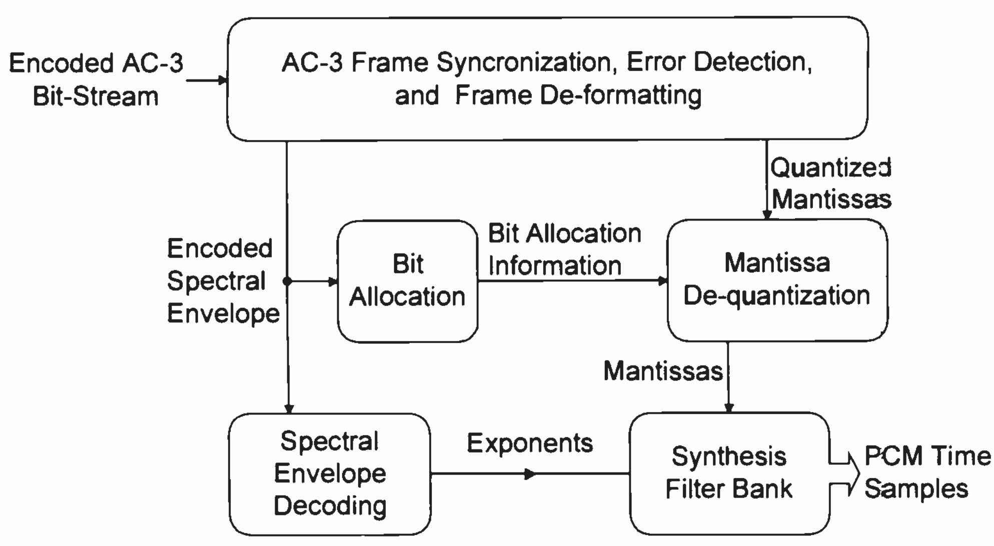


Figure 3.7-10. Overview of the AC-3 audio-compression system decoder.[1]

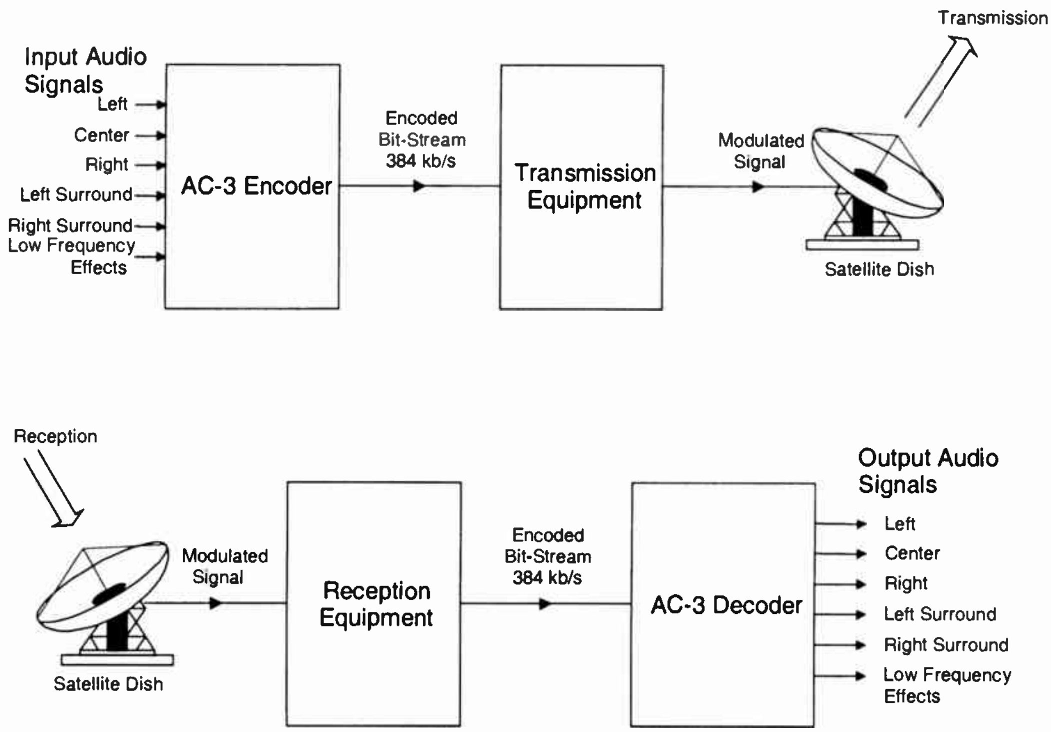


Figure 3.7-8. Example application of the AC-3 audio subsystem for satellite audio transmission.[1]

filterbank. Overlapping blocks of 512 time samples are multiplied by a time window and transformed into the frequency domain. Because of the overlapping blocks, each PCM input sample is represented in two sequential transformed blocks. The frequency domain representation then may be decimated by a factor of 2, so that each block contains 256 frequency coefficients. The individual frequency coefficients are represented in binary exponential notation as a binary exponent and a mantissa. The set of exponents is encoded into a coarse representation of the signal spectrum, referred to as the *spectral envelope*. This spectral envelope is used

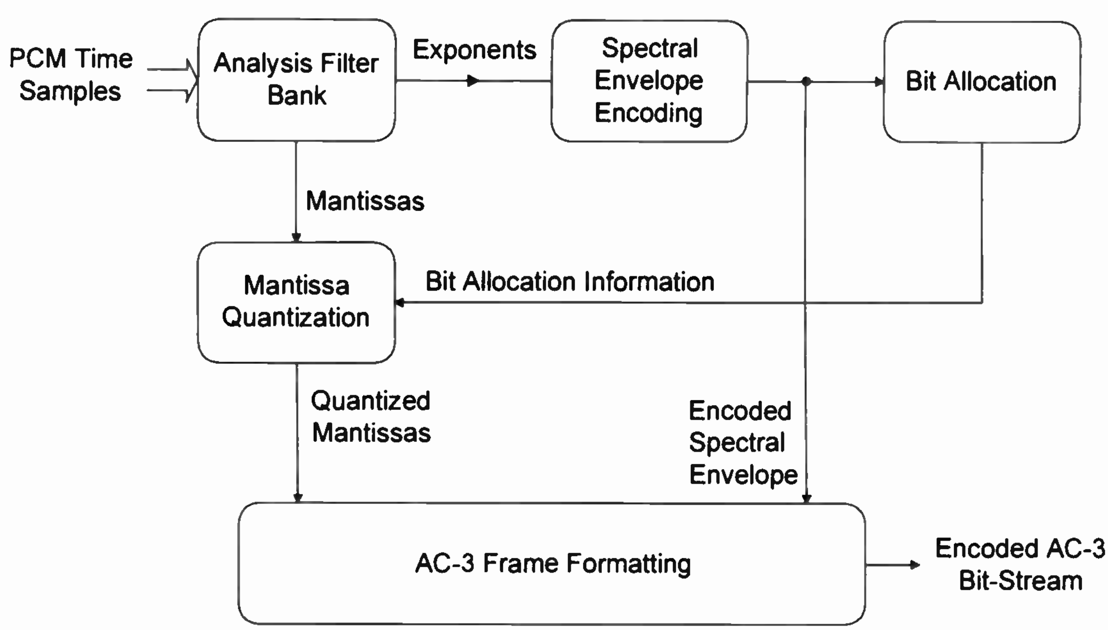


Figure 3.7-9. Overview of the AC-3 audio-compression system encoder.[1]

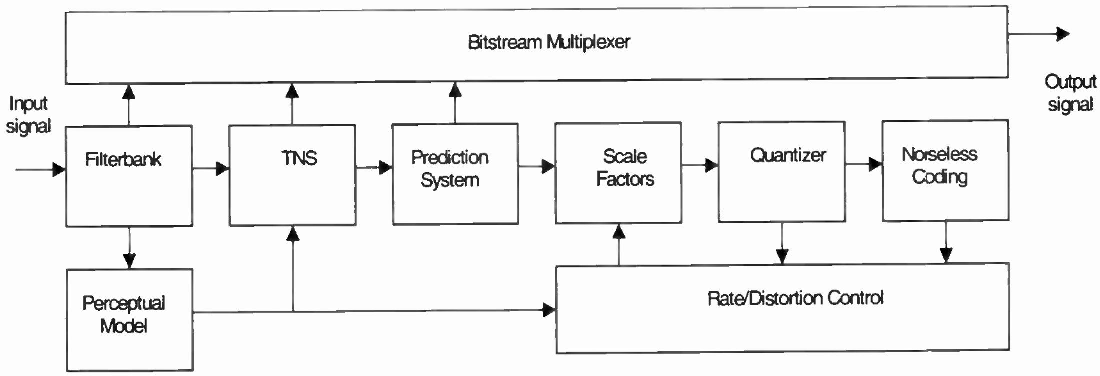


Figure 3.7-7. Functional block diagram of the MPEG-2 AAC coding system.

(maximum). AAC is not backward compatible with the Layer 1 and Layer 2 codes.

MPEG-4

MPEG-4, as with the MPEG-1 and MPEG-2 efforts, is not concerned solely with the development of audio coding standards, but also encompasses video coding and data transmission elements. In addition to building upon the audio coding standards developed for MPEG-2, MPEG-4 includes a revolutionary new element—*synthesized sound.* Tools are provided within MPEG-4 for coding of both natural sounds (speech and music) and for synthesizing sounds based on structured descriptions. The representations used for synthesizing sounds can be formed by text or by instrument descriptions, and by coding other parameters to provide for effects, such as reverberation and spatialization.

Natural audio coding is supported within MPEG-4 at bit rates ranging from 2–64 kbps and includes the MPEG-2 AAC Standard (among others) to provide for general compression of audio in the upper bit rate range (8–64 kbps), the range of most interest to broadcasters. Other types of coders, primarily voice coders (or *vocoders*) are used to support coding down to the 2 kbps rate.

For synthesized sounds, decoders are available that operate based on so-called structured inputs, that is, input signals based on descriptions of sounds and not the sounds themselves. Text files are one example of a structured input. In MPEG-4, text can be converted to speech in a text-to-speech (TTS) decoder. Synthetic music is another example, and may be delivered at extremely low bit rates while still describing an exact sound signal. The standard's structured audio decoder uses a language to define an orchestra made up of instruments, which can be downloaded in the bit stream, not fixed in the decoder.

TTS support is provided in MPEG-4 for unembellished text, or text with prosodic (pitch contour, phoneme duration, etc.) parameters, as an input to generate intelligible synthetic speech. It includes the following functionalities:

- Speech synthesis using the prosody of the original speech
- Facial animation control with phoneme information (important for multimedia applications)
- Trick mode functionality: pause, resume, jump forward, jump backward
- International language support for text
- International symbol support for phonemes
- Support for specifying the age, gender, language and dialect of the speaker.

MPEG-4 does not standardize a method of synthesis, but rather a method of describing synthesis.

AC-3 Audio System for DTV

The AC-3 digital compression algorithm specified in the ATSC DTV system can encode from 1–5.1 channels of source audio from a PCM representation into a serial bit stream at data rates ranging from 32–640 kbps.

A typical application of the bit-reduction algorithm is shown in Figure 3.7-8. In this example, a 5.1 channel audio program is converted from a PCM representation requiring more than 5 Mbps (6 channels $\times$ 48 kHz $\times$ 18 bits = 5.184 Mbps) into a 384 kbps serial bit stream by the AC-3 encoder. Radio frequency (RF) transmission equipment converts this bit stream into a modulated waveform that is applied to a satellite transponder. The amount of bandwidth and power thus required by the transmission has been reduced by more than a factor of 13 by the AC-3 digital compression system. The received signal is demodulated back into the 384 kbps serial bit stream, and decoded by the AC-3 decoder. The result is the original 5.1 channel audio program.

Encoding

The AC-3 encoder accepts PCM audio and produces the encoded bit stream for the ATSC DTV Standard.[1] The AC-3 algorithm achieves high *coding gain* (the ratio of the input bit rate to the output bit rate) by coarsely quantizing a frequency domain representation of the audio signal. A block diagram of this process is given in Figure 3.7-9. The first step in the encoding chain is to transform the representation of audio from a sequence of PCM time samples into a sequence of blocks of frequency coefficients. This is done in the *analysis*

After the number of side information bits required for scale factors, bit allocation codes, cyclic redundancy check (CRC) and other functions have been determined, the remaining bits left in the pool are used in the recoding of the audio subband samples. The allocation of bits for the audio is determined by calculating the SMR, via the FFT, for each of the 12 subband sample blocks. The bit allocation algorithm then selects one of 15 available quantizers with a range such that the overall bit rate limitations are met and the quantization noise is masked as far as possible. If the composition of the audio signal is such that there are not enough bits in the pool to adequately code the subband samples, then the quantizers are adjusted down to a best-fit solution with (hopefully) minimum damage to the decoded audio at the output.

If the signal block being processed lies in the lower one third of the 32 frequency subbands, a 4 bit code word is simultaneously generated to identify the selected quantizer; this word is, again, carried as side information in the main data frame. A 3 bit word would be generated for processing in the mid frequency subbands and a 2 bit word for the higher frequency subbands. When the audio analysis demands it, this allows for at least 15, 7 and 3 quantization levels, respectively, in each of the three spectrum groupings. However, each quantizer can, if required, cover from 3 to 65,535 levels and additionally, if no signal is detected then no quantization takes place.

As with the scale factor data, some further redundancy can be exploited, which increases the efficiency of the quantizing process. For the lowest quantizer ranges (3, 5 and 9 levels), three successive subband sample blocks are grouped into a *granule* and this—in turn—is defined by only one code word. This is particularly effective in the higher frequency subbands where the quantizer ranges are invariably set at the lower end of the scale.

Error detection information can be relayed to the decoder by inserting a 16 bit CRC word in each data frame. This parity check word allows for the detection of up to three single bit errors or a group of errors up to 16 bits in length. A codec incorporating an error concealment regime can either mute the signal in the presence of errors or replace the impaired data with a previous, error free, data frame. The typical data frame structure for ISO/MPEG-1 Layer 2 audio is given in Figure 3.7-6.

MPEG-2 AAC

Also of note is MPEG-2 advanced audio coding (AAC), a highly advanced perceptual code, used initially for digital radio applications. The AAC code improves on previous techniques to increase coding efficiency. For example, an AAC system operating at 96 kbps produces the same sound quality as ISO/MPEG-1 Layer 2 operating at 192 kbps—a 2:1 reduction in bit rate. There are three modes (profiles) in the AAC standard:

- **Main.** Used when processing power, and especially memory, are readily available
- **Low complexity.** (LC) used when processing cycles and memory use are constrained
- **Scaleable sampling rate (SSR).** Appropriate when a *scalable decoder* is required. A scalable decoder can be designed to support different levels of audio quality from a common bit stream; for example, having both high and low cost implementations to support higher and lower audio qualities, respectively

Different profiles trade off encoding complexity for audio quality at a given bit rate. For example, at 128 kbps, the main profile AAC code has a more complex encoder structure than the LC AAC code at the same bit rate, but provides better audio quality as a result.

A block diagram of the AAC system general structure is given in Figure 3.7-7. The blocks in the drawing are referred to as *tools* that the coding algorithm uses to compress the digital audio signal. While many of these tools exist in most audio perceptual codes, two are unique to AAC—the *temporal noise shaper* (TNS) and the *prediction system* tool. The TNS uses a backward adaptive prediction process to remove redundancy between the frequency channels that are created by the prediction system tool.

MPEG-2 AAC provides the capability of up to 48 main audio channels, 16 low frequency effects channels, 16 overdub/multilingual channels, and 10 data streams. By comparison, ISO/MPEG-1 Layer 1 provides two channels and Layer 2 provides 5.1 channels

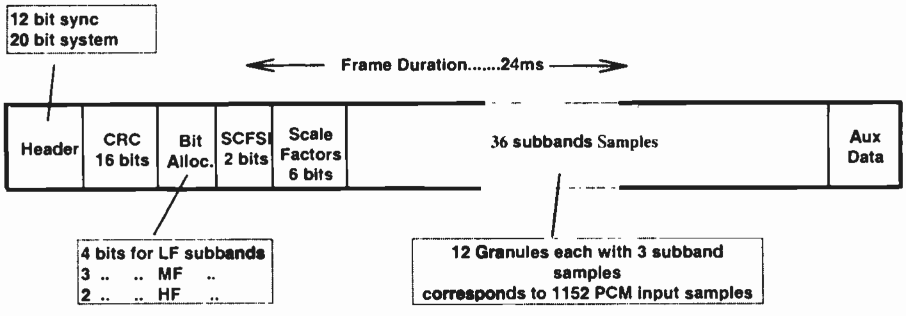
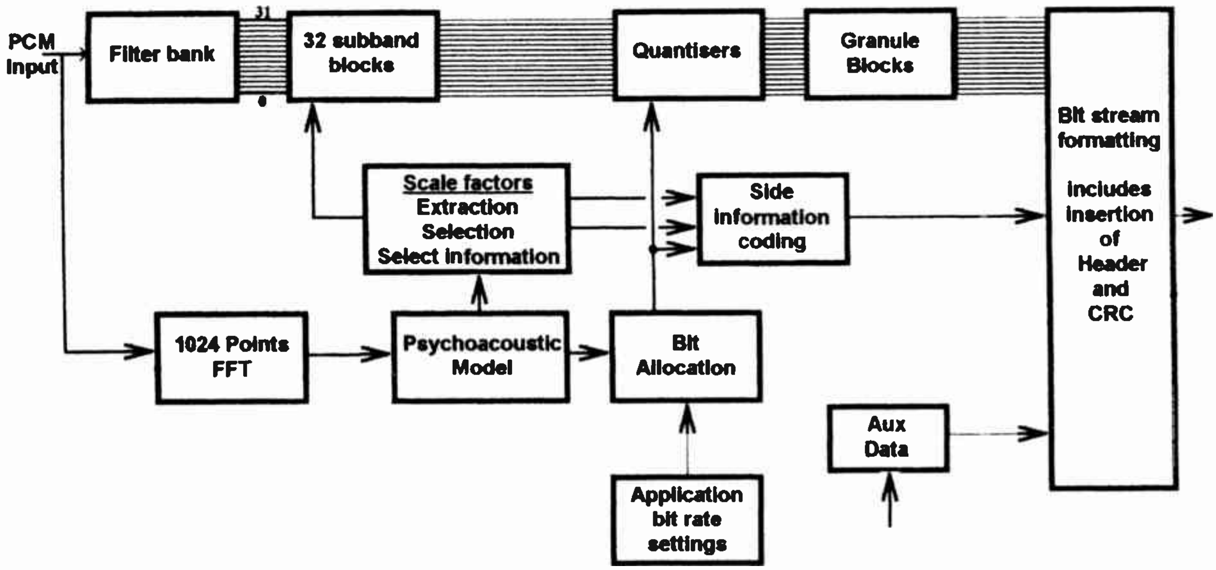


Figure 3.7-6. ISO/MPEG-1 Layer 2 data frame structure.



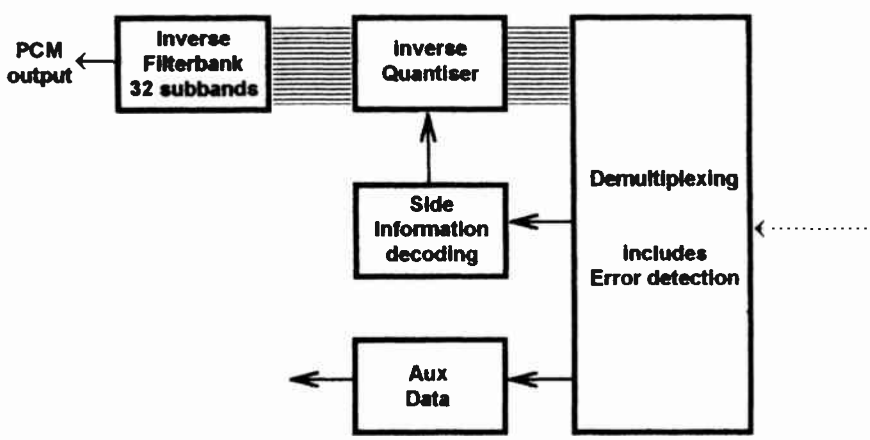


Figure 3.7-5. ISO/MPEG-1 Layer 2 system: (a) encoder block diagram, (b) decoder block diagram.

data rate for these scale factors can be reduced by exploiting some redundancy in the data. Three successive subband scale factors are analyzed and a pattern is determined. This pattern, which is obviously related to the nature of the audio signal, will decide whether one, two or all three scale factors are required. The decision will be communicated by the insertion of an additional *scale factor select information* (SCFSI) data word of 2 bits.

In the case of a fairly stationary tonal-type sound, there will be very little change in the scale factors and only the largest one of the three is transmitted; the corresponding data rate will be (6 + 2) or 8 scale factor bits per frame. However, in a complex sound with rapid changes in content, the transmission of two or even three scale factors may be required, producing a maximum scale factor bit rate demand of (6 + 6 + 6 + 2) or 20 bits per frame. Compared with Layer 1, this method of coding the scale factors reduces the allocation of data bits required for them by half.

The number of data bits allocated to the overall bit pool is limited or fixed by the data rate parameters. These parameters are set out by a combination of sampling frequency, compression ratio, and—where applicable—the transmission medium. In the case of 20 kHz stereo being transmitted over ISDN, for example, the maximum data rate is 384 kbps, sampling at 48kHz, with a compression ratio of 4:1.

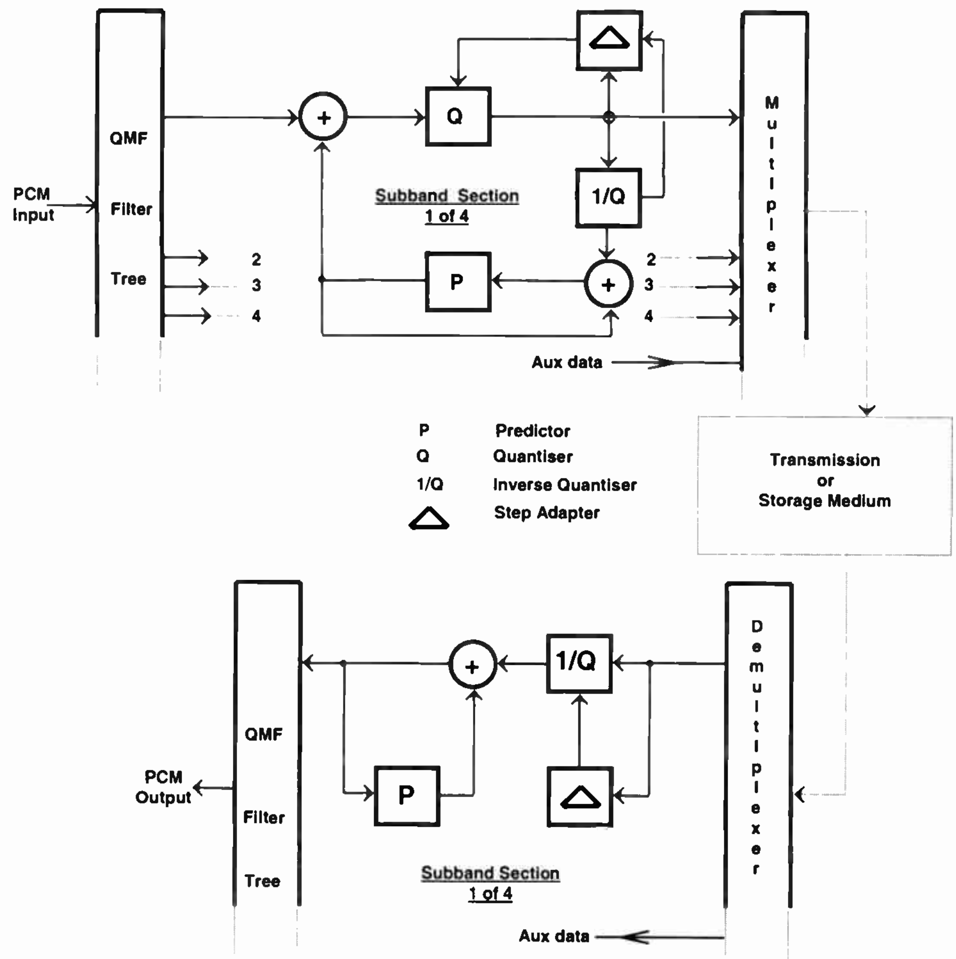


Figure 3.7-4. apt-X100 audio coding system: (a) encoder block diagram, (b) decoder block diagram.

The incoming linear PCM signal block is divided into 32 equally spaced subbands using a polyphase analysis filter bank (see Figure 3.7-5(a)). At 48 kHz sampling, this equates to the bandwidth of each subband being 750 Hz. The bit allocation for the requantizing of these subband samples is then dynamically controlled by information derived from analyzing the audio signal, measured against a preset psychoacoustic model.

The filter bank, which displays manageable delay and minimal complexity, optimally adapts each block of audio to balance between the effects of temporal masking and inaudible pre-echoes.

The PCM signal is also fed to a *fast Fourier transform* (FFT) running in parallel with the filter bank. The aural sensitivities of the human auditory system are exploited by using this FFT process to detect the differences between the wanted and unwanted sounds and the quantization noise already present in the signal, and then to adjust the signal-to-mask thresholds, conforming to a preset perceptual model.

This psychoacoustic model is only found in the coder, thus making the decoder less complex and permitting the freedom to exploit future improvements in coder design. The actual number of levels for each quantizer is determined by the bit allocation. This is arrived at by setting the *signal-to-mask ratio* (SMR) parameter, defined as the difference between the minimum masking threshold and the maximum signal level. This minimum masking threshold is calculated using the psychoacoustic model and provides a reference noise level of *just noticeable* noise for each subband.

In the decoder, after demultiplexing and deciphering of the audio and side information data, a dual synthesis filter bank reconstructs the linear PCM signal in blocks of 32 output samples (see Figure 3.7-5(b)).

A scale factor is determined for each 12 subband sample block. The maximum of the absolute values of these 12 samples generates a scale factor word consisting of 6 bits, a range of 63 different levels. Because each frame of audio data in Layer 2 corresponds to 36 subband samples, this process will generate 3 scale factors per frame. However, the transmitted

- **The ATSC DTV system.** Uses the subband APCM algorithm in Dolby AC-3 for the audio surround system associated with the ATSC DTV Standard. AC-3 delivers five audio channels plus a bass-only effects channel in less bandwidth than that required for one stereo CD channel. This configuration is referred to as 5.1 channels.

For the purposes of illustration, two commonly used audio compression systems will be examined in some detail:

- apt-X100
- ISO/MPEG-1 Layer 2

apt-X100

apt-X100 is a four subband prediction (ADPCM) algorithm. Differential coding reduces the bit rate by coding and transmitting or storing only the difference between a predicted level for a PCM audio sample and the absolute level of that sample, thus exploiting the redundancy contained in the PCM signal.

Audio exhibits relatively slowly varying energy fluctuations with respect to time. Adaptive differential coding, which is dependent on the energy of the input signal, dynamically alters the step size for each quantizing interval to reflect these fluctuations. In apt-X100, this equates to a backward adaptation process and involves the analysis of 122 previous samples. Being a continuous process, this provides an almost constant and optimal signal-to-quantization noise ratio across the operating range of the quantizer.

Time domain subband algorithms implicitly model the hearing process and indirectly exploit a degree of irrelevancy by accepting that the human ear is more sensitive at lower frequencies. This is achieved in the four subband derivative by allocating more bits to the lower frequency bands. This is the only application of psychoacoustics exercised in apt-X100. All the information contained in the PCM signal is processed, audible or not (no attempt is made to remove irrelevant information). It is the unique fixed allocation of bits to each of the four subbands, coupled with the filtering characteristics of each individual listener's hearing system, that achieves the satisfactory audible end result.

The user-defined output bit rate ranges from 56–384 kbps, achieved by using various sampling frequencies from 16–48 kHz, which produce audio bandwidths from 7.5 kHz mono to 22 kHz stereo.

Auxiliary data up to 9.6 kbps can also be embedded into the data stream without incurring a bit overhead penalty. When this function is enabled, an audio bit in one of the higher frequency subbands is replaced by an auxiliary data bit, again with no audible effect.

An important feature of this algorithm is its inherent robustness to random bit errors. No audible distortion is apparent for normal program material at a bit error rate (BER) of 1:10,000 (1×10^{-4}), while speech is still intelligible down to a BER of 1:10 (1×10^{-1}).

Distortions introduced by bit errors are constrained within each subband and their impact on the decoder subband predictors and quantizers is proportional to the magnitude of the differential signal being decoded at that instant. Thus, if the signal is small—which will be the case for a low level input signal or for a resonant, highly predictable input signal—any bit error will have minimal effect on either the predictor or quantizer.

The 16 bit linear PCM signal is processed in time blocks of four samples at a time. The 64 bits are filtered into four equal width frequency subbands—for 20 kHz, this would be 0–5 kHz, 5–10 kHz and so on. The four outputs from the quadrature mirror filter (QMF) tree are still in the 16 bit linear PCM format, but are now frequency limited.

As shown in Figure 3.7-4, the compression process can be mapped by taking, for example, the first and lowest frequency subband. The first step is to create the difference signal. After the system has settled down on initiation, there will be a reconstructed 16 bit difference signal at the output of the inverse quantizer. This passes into a prediction loop that, having analyzed 122 previous samples, will make a prediction for the level of the next full level sample arriving from the filter tree. This prediction is then compared with the actual level.

The output of the comparator is the resulting 16 bit difference signal. This is requantized to a new 7 bit format, which in turn is inverse quantized back to 16 bits again to enable the prediction loop.

The output from the inverse quantizer is also analyzed for energy content, again for the same 122 previous samples. This information is compared with onboard look up tables and a decision is made to dynamically adjust, up or down as required, the level of each step of the 1024 intervals in the 7 bit quantizer. This ensures that the quantizer will always have adequate range to deal with the varying energy levels of the audio signal. Therefore, the input to the multiplexer will be a 7 bit word but the range of those bits will be varying in relation to the signal energy.

The three other subbands will go through the same process, but the number of bits allocated to the quantizers are much less than for the first subband.

The output of the multiplexer or bit stream formatter is a new 16 bit word that represents four input PCM samples and is, therefore, one quarter of the input rate—a reduction of 4:1.

The decoding process is the complete opposite of the coding procedure. The incoming 16 bit compressed data word is demultiplexed and used to control the operation of four subband decoder sections, each with similar predictor and quantizer step adjusters. A QMF filter tree finally reconstructs a linear PCM signal and separates any auxiliary data that may be present.

ISO/MPEG-1 Layer 2

This algorithm differs from MPEG Layer 1 by adopting more accurate quantizing procedures and removing any apparent redundancy from the generated scale factors. The ISO/MPEG-1 Layer 2 scheme operates on a block of 1152 PCM samples, which at 48 kHz sampling represents a 24 ms time block of the input audio signal. Simplified block diagrams of the encoding/decoding systems are given in Figure 3.7-5.

several factors tend to limit the number of subbands employed. The primary considerations include:

- The level variation of normal audio signals leading to an averaging of the energy across bands and a subsequent reduction in the coding gain
- The coding or processing delay introduced by additional subbands
- The overall computational complexity of the system

The two key issues in the analysis of a subband framework are:

- Determining the likely improvement associated with additional subbands
- Determining the relationships between subband gain, the number of subbands, and the response of the filter bank used to create those subbands

APCM Coding

The APCM processor acts in a similar fashion to an automatic gain control system, continually making adjustments in response to the dynamics—at all frequencies—of the incoming audio signal. Transform coding takes a time block of signal, analyzes it for frequency and energy and identifies irrelevant content. Again, to exploit the spectral response of the ear, the frequency spectrum of the signal is divided into a number of subbands, and the most important criteria are coded with a bias toward the more sensitive low frequencies. At the same time, through the use of psychoacoustic masking techniques, those frequencies which it is assumed will be masked by the ear are also identified and removed. The data generated, therefore, describes the frequency content and the energy level at those frequencies, with more bits being allocated to the higher energy frequencies than those with lower energy.

The larger the time block of signal being analyzed, the better the frequency resolution and the greater the amount of irrelevancy identified. The penalty, however, is an increase in coding delay and a decrease in temporal resolution. A balance has been struck with advances in perceptual coding techniques and psychoacoustic modeling leading to increased efficiency. It is reported that, with this approach to compression, some 80% of the input audio can be removed with acceptable results.

This hybrid arrangement of working with time domain subbands and simultaneously carrying out a spectral analysis can be achieved by using a dynamic bit allocation process for each subband. This subband APCM approach is found in the popular range of software-based MUSICAM, Dolby AC-2 and ISO/MPEG-1 Layers 1 and 2 algorithms. Layer 3—a more complex method of coding and operating at much lower bit rates—is, in essence, a combination of the best functions of MUSICAM and ASPEC, another adaptive transform algorithm. Table 3.7-1 lists the primary operational parameters for these systems.

Additionally, some of these systems exploit the significant redundancy between stereo channels by using a technique known as joint stereo coding. After the common information between left and right channels of a stereo signal has been identified, it is coded only once, thus reducing the bit rate demands yet again.

Each of the subbands has its own defined masking threshold. The output data from each of the filtered subbands is requantized with just enough bit resolution to maintain adequate headroom between the quantization noise and the masking threshold for each band. In more complex coders (ISO/MPEG-1 Layer 3), any spare bit capacity is utilized by those subbands with the greater need for increased masking threshold separation. The maintenance of these signal-to-masking threshold ratios is crucial if further compression is contemplated for any postproduction or transmission process.

Common Compression Techniques

Subband APCM coding has found numerous applications in the professional audio industry, including:

- **The DCC.** Uses the simplest implementation of subband APCM with the PASC/ISO/MPEG-1 Layer 1 algorithm incorporating 32 subbands offering 4:1 compression and producing a bit rate of 384 kbps
- **The MiniDisc with the proprietary ATRAC algorithm.** Produces 5:1 compression and 292 kbps bit rate. This algorithm uses a modified discrete cosine transform (MDCT) technique ensuring greater signal analysis by processing time blocks of the signal in non uniform frequency divisions, with fewer divisions being allocated to the least sensitive higher frequencies
- **ISO/MPEG-1 Layer 2 (MUSICAM).** A software-based algorithm that can be implemented to produce a range of bit rates and compression ratios commencing at 4:1

Table 3.7-1
Operational parameters of subband APCM algorithm.

Coding System	Compression Ratio	Subbands	Bit Rate, kbits/s	A to A Delay, ms[1]	Audio Bandwidth, kHz
Dolby AC-2	6:1	256	256	45	20
ISO Layer 1	4:1	32	384	19	20
ISO Layer 2	Variable	32	192–256	>40	20
ISO Layer 3	12:1	576	128	>80	20
MUSICAM	Variable	32	128–384	>35	20

[1] The total system delay (encoder to decoder) of the coding system.

both the noise floor and the audio bandwidth will be set by the stage operating at the lowest bit rate. It is, therefore, worth emphasizing that after these platforms have been set by a low bit rate stage, they cannot be subsequently improved by using a following stage operating at a higher bit rate.

A stage of compression may well be followed in the audio chain by another digital stage, either of compression or linear, but—more importantly—operating at a different sampling frequency. If a D/A conversion is to be avoided, a sample rate converter must be used. This can be a stand-alone unit or it may already be installed as a module in existing equipment. Where a following stage of compression is operating at the same sampling frequency but a different compression ratio, the bit resolution will change by default.

If the stages have the same sampling frequencies, a direct PCM or AES/EBU digital link can be made, thus avoiding the need for sample rate or D/A conversion.

Editing Compressed Data

The resolution of the compressed data may or may not be adequate to allow direct editing of the audio signal. The minimum audio sample that can be removed or edited from a transform-coded signal will be determined by the size of the time block of the PCM signal being analyzed. The larger the time block the more difficult the editing of the compressed data becomes.

Prediction and Transform Algorithms

Most audio compression systems are based upon one of two basic technologies:

- Predictive or adaptive differential PCM (ADPCM) time domain coding
- Transform or adaptive PCM (APCM) frequency domain coding

It is in their approaches to dealing with the redundancy and irrelevancy of the PCM signal that these techniques differ.

The time domain or prediction approach includes G.722, which has been a universal standard since the mid 70s, and was joined in 1989 by a proprietary algorithm, apt-X100. Both these algorithms deal mainly with redundancy.

The frequency domain or transform method adopted by a number of algorithms deal in irrelevancy, adopting psychoacoustic masking techniques to identify and remove those unwanted sounds. This range of algorithms includes the industry standards ISO/MPEG-1 and 2; Layers 1, 2, and 3; MPEG AAC; MUSICAM; Dolby AC-2 and AC3, apt-Q and others.

Subband Coding

Without exception all of the algorithms just mentioned process the PCM signal by splitting it into a number of frequency subbands, in one case as few as two (G.722) or as many as 1024 (apt-Q). MPEG-1 Layer 1, with 4:1 compression, has 32 frequency subbands and is the system found in the Digital Compact Cassette (DCC). The MiniDisc ATRAC proprietary algorithm at 5:1 has a more flexible multisubband approach, which is dependent on the complexity of the audio signal.

Subband coding enables the frequency domain redundancies within the audio signals to be exploited. This permits a reduction in the coded bit rate, compared to PCM, for a given signal fidelity. Spectral redundancies are also present as a result of the signal energies in the various frequency bands being unequal at any instant in time. By altering the bit allocation for each subband, either by dynamically adapting it according to the energy of the contained signal or by fixing it for each subband, the quantization noise can be reduced across all bands. This process compares favorably with the noise characteristics of a PCM coder performing at the same overall bit rate.

On its own, subband coding, incorporating PCM in each band, is capable of providing a performance improvement or gain compared with that of full band PCM coding, both being fed with the same complex, constant level input signal. The improvement is defined as subband gain and is the ratio of the variations in quantization errors generated in each case while both are operating at the same transmission rate. The gain increases as the number of subbands increase, and with the complexity of the input signal. However, the implementation of the algorithm also becomes more difficult and complex.

Quantization noise generated during the coding process is constrained within each subband and cannot interfere with any other band. The advantage of this approach is that the masking by each subbands dominant signals is much more effective because of the reduction in the noise bandwidth. Figure 3.7-3 charts subband gain as a function of the number of subbands for four essentially stationary, but differing, complex audio signals.

In practical implementations of compression codecs,

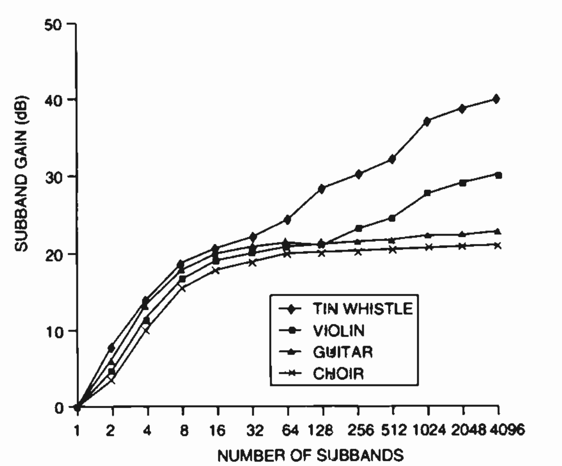


Figure 3.7-3. Variation of subband gain as a function of the number of subbands.

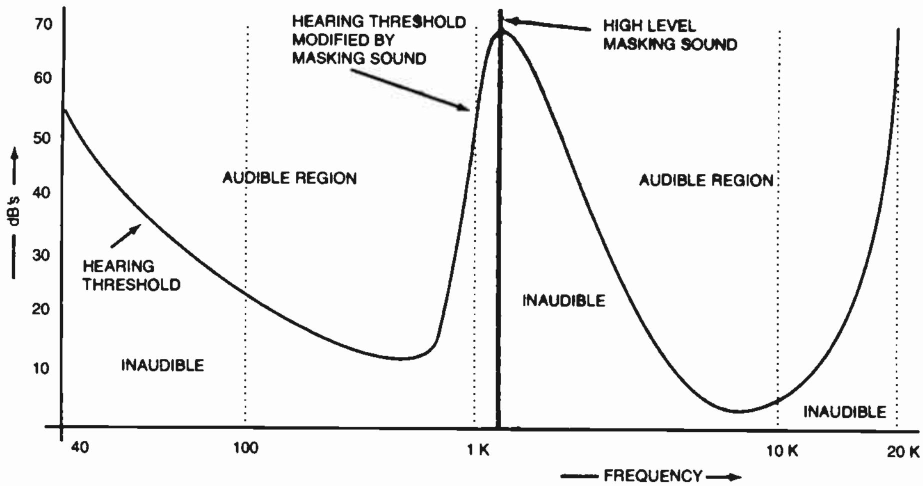


Figure 3.7-2. Example of the masking effect of a high-level sound.

Compression or bit rate reduction of the PCM signal leads to the requantizing of an already quantized signal, which will unavoidably inject further quantization noise. It always has been good operating practice to restrict the number of A/D and D/A conversions in an audio chain. Nothing has changed in this regard, and now the number of compression stages also should be kept to a minimum. Additionally, the bit rates of these stages should be set as high as practical or, the compression ratio should be as low as possible.

Sooner or later—after a finite number of A/D, D/A conversions and passes of compression coding, of whatever type—the accumulation of quantization noise and other unpredictable signal degradations eventually will break through the noise/signal threshold, be interpreted and processed as part of the audio signal and heard by the listener.

Sampling Frequency and Bit Rate

The bit rate of a digital signal is defined by:

sampling frequency × *bit resolution* × *number of audio channels*

The rules regarding the selection of a sampling frequency are based on *Nyquist's theorem.* This ensures that, in particular, the lower sideband *alias,* with its dc component located at the sampling frequency, does not encroach into the baseband audio. Objectionable and audible aliasing effects would occur if the two bands were to overlap. In practice, the sampling rate is set slightly above twice the highest audible frequency, which makes the necessary filter designs less complex and less expensive.

In the case of a stereo CD with the audio signal having been sampled at 44.1 kHz, this sampling rate produces audio bandwidths of approximately 20 kHz for each channel. The resulting audio bit rate = 44.1 kHz × 16 × 2 = 1.411 Mbps, as discussed previously.

Processing and Propagation Delay

As noted previously, the current range of popular compression algorithms operate—for all intents and purposes—in real-time. However, this process does of necessity introduce some measurable delay into the audio chain. All algorithms take a finite time to analyze the incoming signal, which can range from a few milliseconds to tens and even hundreds of milliseconds. The amount of processing delay will be crucial if the equipment is to be used in any interactive or two-way application. As a rule of thumb, any more than 20 ms of delay in a two way audio exchange is problematic. Propagation delay in satellite and long terrestrial circuits is a fact of life. A two-way hook up over a 1000 km, full duplex, telecom digital link has a propagation delay of 3 ms in each direction. This is comparable to having a conversation with someone standing 1 m away. It is obvious that even over a very short distance, the use of a codec with a long processing delay characteristic will have a dramatic effect on operation.

Bit Rate and Compression Ratio

The ITU has recommend the following bit rates when incorporating data compression in an audio chain:

- 128 kbps per mono channel (256 kbps for stereo) as the minimum bit rate for any stage if further compression is anticipated or required
- 192 kbps per mono channel (384 kbps for stereo) as the minimum bit rate for the first stage of compression in a complex audio chain.

These markers place a 4:1 compression ratio at the safe end in the scale. However, more aggressive compression ratios, currently up to a nominal 20:1, are available. Keep in mind, though, that low bit rate, high level compression can lead to problems if any further stages of compression are required or anticipated.

With successive stages of compression, either or that is highly predictable and, therefore, can be considered redundant.

Redundancy, measurable and quantifiable, can be removed in the coder and replaced in the decoder; this process often is referred to as statistical compression. *Irrelevancy,* on the other hand, referred to as perceptual coding, once removed from the signal cannot be replaced and is lost, irretrievably. This is entirely a subjective process, with each proprietary algorithm using a different psychoacoustic model.

Critically perceived signals, such as pure tones, are high in redundancy and low in irrelevancy. They compress quite easily, almost totally a statistical compression process. Conversely, noncritically perceived signals, such as complex audio or noisy signals, are low in redundancy and high in irrelevancy. These compress easily in the perceptual coder, but with the total loss of all the irrelevant content.

Human Auditory System

The sensitivity of the human ear is biased toward the lower end of the audible frequency spectrum, around 3 kHz. At 50 Hz, the bottom end of the spectrum, and 17 kHz at the top end, the sensitivity of the ear is down by approximately 50 dB relative to its sensitivity at 3 kHz (see Figure 3.7-1). Additionally, very few audio signals—music or speech based—carry fundamental frequencies above 4 kHz. Taking advantage of these characteristics of the ear, the structure of audible sounds, and the redundancy content of the PCM signal is the basis used by the designers of the *predictive* range of compression algorithms.

Another well-known feature of the hearing process is that loud sounds mask out quieter sounds at a similar or nearby frequency. This compares with the action of an automatic gain control, turning the gain down when subjected to loud sounds, thus making quieter sounds less likely to be heard. For example, as illustrated in Figure 3.7-2, if we assume a 1 kHz tone at a level of 70 dBu, levels of greater than 40 dBu at 750 Hz and 2 kHz would be required for those frequencies to be heard. The ear also exercises a degree of temporal masking, where a strong sound will mask a weaker sound which is close by, before or after, in time proximity.

It is by mimicking these additional psychoacoustic features of the human ear and identifying the irrelevancy content of the input signal that the *transform* range of low bit rate algorithms operate, adopting the principle that if the ear is unable to hear the sound then there is no point in transmitting it in the first place.

Quantization

Quantization is the process of converting an analog signal to its representative digital format or, as in the case with compression, the requantizing of an already converted signal. This process is the limiting of a finite level measurement of a signal sample to a specific preset integer value. This means that the *actual* level of the sample may be greater or smaller than the preset *reference* level it is being compared with. The difference between these two levels, called the *quantization error,* is compounded in the decoded signal as *quantization noise.*

Quantization noise, therefore, will be injected into the audio signal after each analog-to-digital (A/D) and digital-to-analog (D/A) conversion, the level of that noise being governed by the bit allocation associated with the coding process (the number of bits allocated to represent the level of each sample taken of the analog signal). For linear PCM, the bit allocation is commonly 16. The level of each audio sample, therefore, will be compared with one of 2^{16}, or 65,536 discrete levels or steps.

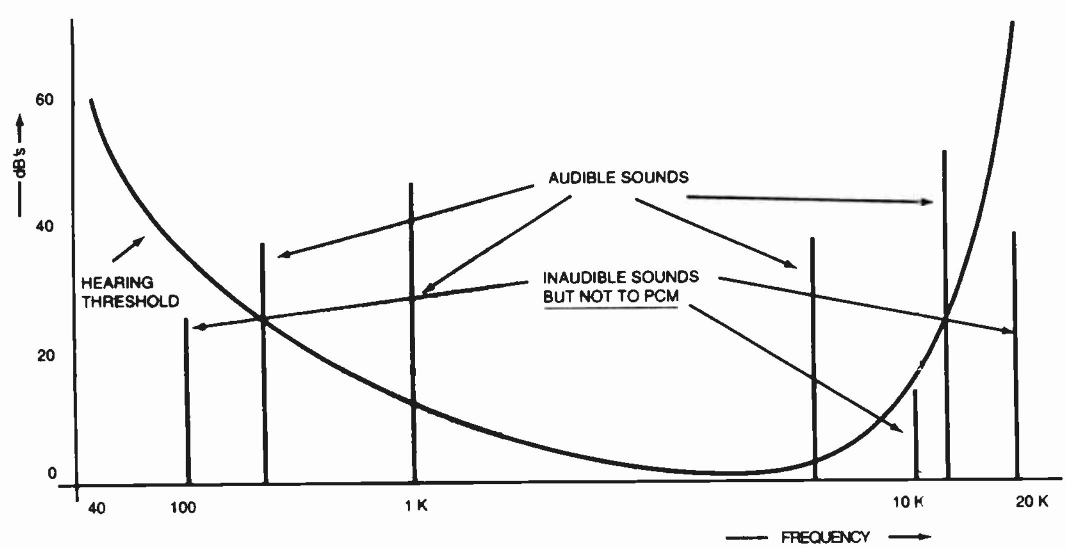


Figure 3.7-1. Generalized frequency response of the human ear. Note how the PCM process captures signals that the ear cannot distinguish.

3.7 DIGITAL AUDIO COMPRESSION TECHNOLOGIES

FRED WYLIE
AUDIO PROCESSING TECHNOLOGY, LTD.
BELFAST, NORTHERN IRELAND

INTRODUCTION

Virtually all applications of digital audio deal with an enormous amount of data. Because of this, compression is an integral part of most modern digital audio systems. In fact, compression is essential to the ATSC DTV and digital audio broadcasting (DAB) systems.

A number of existing and proposed audio-compression systems employ a combination of processing techniques. Any scheme that becomes widely adopted can enjoy economies of scale and reduced market confusion. Timing, however, is critical to market acceptance of any standard. If a standard is selected well ahead of market demand, more cost-effective or higher-performance approaches may become available before the market takes off. On the other hand, a standard may be merely academic if it is established after alternative schemes already have become well entrenched in the marketplace.

These forces are shaping audio technology of the future. Any number of scenarios have been postulated as to the hardware and software that will drive the digital audio facility of the year 2000 and beyond. One thing is certain, however: It will revolve around compressed signals.

As with video, high on the list of priorities for the professional audio industry is to refine and extend the range of digital equipment capable of the capture, storage, post production, exchange, distribution and transmission of high quality audio, be it mono, stereo, or multichannel surround sound. This demand is being driven by end-users, broadcasters, film makers and the recording industry alike, who are moving rapidly towards a tapeless environment. Over the last two decades, there have been continuing advances in *digital signal processing* (DSP) technology, which have supported research engineers in their endeavors to produce the necessary hardware, particularly in the field of digital audio data compression or—as it is often referred to—*bit rate reduction.* There exist a number of real-time or—in reality—near instantaneous compression coding algorithms. These can significantly lower the circuit bandwidth and storage requirements for the transmission, distribution and exchange of high-quality audio.

The introduction in 1983 of the compact disc (CD) digital audio format set a quality benchmark that the manufacturers of subsequent professional audio equipment strive to match or improve. The discerning consumer now expects the same quality from radio and television receivers. This leaves the broadcaster with an enormous challenge.

It can be an expensive and complex technical exercise to fully implement a linear *pulse code modulation* (PCM) infrastructure, except over very short distances and within studio areas. To demonstrate the advantages of distributing compressed digital audio over wireless or wired systems and networks, consider again the CD format as a reference. The CD is a 16 bit linear PCM process, but it has one major handicap: the amount of circuit bandwidth the digital signal occupies in a transmission system. A stereo CD transfers data at 1.411 Mbps, which would require a circuit with a bandwidth of approximately 700 kHz to avoid distortion of the digital signal. In practice, additional bits are added to the signal for channel coding, synchronization, and error correction; this increases the bandwidth demands yet again. 1.5 MHz is the commonly quoted bandwidth figure for a circuit capable of carrying a CD or similarly coded linear PCM digital stereo signal. This can be compared with the 20 kHz needed for each of two circuits to distribute the same stereo audio in the analog format, a 75 fold increase in bandwidth requirements.

AUDIO BIT RATE REDUCTION

In general, analog audio transmission requires fixed input and output bandwidths. This condition implies that in a real-time compression system, the quality, bandwidth and distortion/noise level of both the original and the decoded output sound should not be subjectively different, thus giving the appearance of a lossless and real-time process.

In a technical sense, all real-time bit rate reduction systems can be referred to as *lossy.* In other words, the digital audio signal at the output is not identical to the input signal data stream. However, some compression algorithms are, for all intents and purposes, lossless—they lose as little as 2% of the original signal. Others remove approximately 80% of the original signal.

Redundancy and Irrelevancy

A complex audio signal contains a great deal of information, some of which, because the human ear cannot hear it, is deemed irrelevant. The same signal, depending on its complexity, also contains information

RAID 4 or 5 but lacked the control features required to fine tune the RAID for demanding multi-media applications. However, the current generation of these devices provides outstanding performance and includes a complete set of sophisticated control and maintenance features. Functions such as multiple RAID Level support, choice of cache write policy and battery backup options, on-line spare drive support, dynamically adjustable rebuild rates, and drive enclosure monitoring and alarm support are common. And since the controller is a part of the server PC, the interface to monitor detailed operational statistics and access the administration utilities runs directly from the server console—no other PC is required. In addition, some of these controllers include software that will send the user a page, e-mail or a fax with an error code if the system encounters a failure, and will allow a system administrator to monitor and perform RAID maintenance from a remote PC. And best of all, these devices are becoming more affordable due to the economy of scale associated with the PCI bus and the chipsets involved.

It is very important that the data stored on a centralized server be secure and a redundant disk sub system is the best way to ensure this. Whether you decide to use standard duplexing or RAID technology depends mostly on your comfort level with the setup and maintenance of the technology.

SUMMARY

Station automation is a continuously evolving entity, especially now that a larger portion of it emanates from software running on generic data processing "bit boxes." No matter when you decide to join the digital age, advancements are occurring so quickly that what is rocket science today will surely be obsolete before it wears out. It is therefore wise to use the most stable, reliable advanced technology available today and use a platform that is off the shelf so that it can be continuously upgraded with emerging technology.

Bibliography

Bird, Robert, "Hard Disk Recording for Broadcast Use," *Broadcast Engineering*, Overland Park, KS, pg. 60, August 1991.

Cabot, Richard, "Performance Aspects of Digital Oversampling," *Broadcast Engineering*, Overland Park, KS, pp. 26–93, July 1990

Carlson, Bruce, *Communication Systems*, 2nd Edition, McGraw-Hill, Inc., New York, NY, pp. 294–326, 1975.

Carter, R. Scott and Stephanae Ann Stevenson, "Strike Up the Bandwidth," *NetWare Connection*, Orem, UT, pp. 17–27, July/August 1994.

Conover, Joel, "ATM Fast, Fast Ethernet, Fibre Channel Fast," *Network Computing*, Manhasset, NY, pp. 46–60, November 1995.

ENCO Systems, *DAD_{PRO} Digital Audio Delivery System Reference Manual Revision 2.0*, ENCO Systems, Farmington Hills, MI, pp. 12-1–12-26, 1997.

Heywood, Drew, CNE, Dulaney, E., Homer, B., Niedermiller-Chaffins, D., Orr, S., Stevens, S., and Stone, H., *Inside NetWare 3.12*, Fifth Edition, New Riders Publishing, Indianapolis, IN, pp. 61–797, 1995.

Katron Technologies Inc., *Installation Guide: KTI 10/100VG LAN Adapters*, Katron Technologies Inc., Miami, FL, pp. 1–38, 1994.

Katron Technologies Inc., *Installation Guide: Fast Ethernet PCI Adapter*, Katron Technologies Inc., Miami, FL, pp. 1–19, 1996.

Lamoray, Larry, "Network Basics," Letter from DAD, Vol. 1, No. 2, ENCO Systems, Farmington Hills, MI, pg. 3, May 1994.

Pizzi, Skip, "Digital Audio Workstations Diversify," *Broadcast Engineering*, Overland Park, KS, pp. 56–66, August 1991.

Pohlmann, Ken, *Principles of Digital Audio*, 2nd Edition, SAMS, Carmel, IN, pp.41–98, 1989.

Rodgers, Adam, "Understanding RAID Technology," *Inside NetWare* , Vol. 1, No. 9, The Cobb Group, Louisville, KY, September 1992.

Smyth, Stephen, "Digital Audio Data Compression," *Broadcast Engineering*, Overland Park, KS, February 1992.

Storage Computer Corporation, "RAID Aid: A Taxonomic Extension of the Berkeley Disk Array Schema," Storage Computer Corporation, Nashua, NH, pp. 1–4, 1991.

Whitmann, Art, "Hey, Hewlett-Packard! Let's Stop The Insanity!" *Network Computing*, Manhasset, NY, pp. 103–105, May 1996.

Level Number	Name	Description	Minimum No. Drives	Fault Tolerance
0	Striping (see Spanning in table above)	Block striping is available. Level 0 yields higher performance than is possible with individual drives. I/O transfer speed is increased. No data redundancy: loss of a single component can lead to unrecoverable data loss.	2	No
1	Disk Mirroring	Drives are paired and mirrored, resulting in twice the cost per megabyte (MB). All data is 100 percent duplicated on the mirrored drive.	2	Yes
0+1	Striping and Mirroring	Combination of RAID levels 0 and 1. Data is striped across several physical drives. Each drive has a mirror drive. This level provides redundancy through mirroring.	3	Yes
2	Disk Array	Hamming code error checks across the disks. Data can be recovered without complete duplication of all data, although several check disks are needed. All disks in a group must be accessed, even for transfers, and the slowest must finish before the transfer is complete.		Yes
3	Disk Array	A single parity drive accomplishes redundancy. Parity information is interleaved at the byte level. Drive spindles are typically synchronized; all disks must be accessed and the slowest must complete the process before a transfer is complete.	9	Yes
4	Disk Array (Striping with Fixed Parity)	Interleaving parity at the sector or transfer level results in faster individual disk reads for small transfers. The parity check disk becomes a throughput bottleneck.	3	Yes
5	Disk Array (Striping with Floating Parity)	Parity information is spiraled across all data drives, which combats the throughput bottleneck problem in Level 4. Distributed parity increases write performance but introduces high overhead to track the location of parity addresses.	3	Yes
JBOD	Just a Bunch of Drives	Each drive is operated independently like a normal disk controller, or multiple drives can be spanned and seen as a single large drive. No data redundancy.	1	No

Figure 3.6-15. RAID (Redundant Array of Independent Disks) features.

extensive write back caching and provide battery backup options to protect unflushed data. Most of these devices include extensive control/setup programs that allow the user to fine tune critical parameters such as the level of resources devoted to rebuild a failed/replaced drive. Additional features, like audible alarms to indicate trouble and on-line spare drives that are automatically rebuilt in the event of failure, make these units fairly easy to keep up and running. However, only a few of the more expensive units provide a standard monitor and keyboard interface to their control utilities. Most of these units must be monitored and controlled through a serial port using a serial communications program running on a separate PC or via a limited built-in keypad and cryptic LCD display panel.

A controller based RAID uses a specialized multi-channel disk controller card that plugs into an expansion slot on the host PC (server). These controllers typically contain on-board processing power and memory that is used to administer the RAID functions. Early versions of these RAID controllers often had enough horsepower and memory to run an effective

Disk Configuration	Description, advantages, and disadvantages
Non-Redundant Disks with Tape backup	Backup software is used to save data to a data tape, which is then safely stored. *Advantages* • Good for recovering from catastrophic failure *Disadvantages* • Data transfer is relatively slow compared to a disk drive. You must exit applications and spend time on the backup process • Snapshot of data at time of backup-backup data is only as current as the most recent backup • Data must be accessed linearly and sequentially, so restoring data can also be time-consuming
Disk Mirroring	Identical disk drive is used as backup device. Data on first drive is automatically replicated onto second drive, either by an operating system (like Novell NetWare) or by a special controller (as with RAID technology). *Advantages* • Network operation is unaffected by the failure of one drive-the network just uses the other drive • You can use more than two disks in order to increase redundancy even further • System performance during disk reads may be enhanced, because the system or controller can retrieve the data from whichever disk is better-positioned to provide the data *Disadvantages* • Disk size requirements for data are doubled
Disk Duplexing	Similar to Disk Mirroring above, only the primary and backup drives use separate controllers and cables. This provides another level of hardware fault tolerance. *Advantages* • In most cases, the system will be unaffected by a failure of any component of the subsystem, including drive, controller, and cable • Additional read performance benefit over Disk Mirroring, because additional data buses are now available to transmit data • Write performance advantage over Disk Mirroring, because multiple controllers are available to cache and write data to drives simultaneously *Disadvantages* • Same as Disk Mirroring
Spanning (Striping)	Data is divided between two or more drives. Operating system or controller divides data between drives. Total capacity equals the sum of all drives. *Advantages* • Disk I/O throughput is increased by a factor equal to the number of drives *Disadvantages* • If one of the drives fails, *all data on all drives is lost*–and since total number of components is increased, probability of failure is increased • Offers no redundancy
Mirroring/ Duplexing + Spanning (Striping)	Uses a set of spanned drives; each drive has a mirrored drive. *Advantages* • Increased performance and additive capacity of spanning (striping) • Redundancy and performance advantages of mirroring/duplexing • Failure of single component often undetectable with respect to overall system performance *Disadvantages* • Must routinely monitor components, since failure is sometimes difficult to detect with respect to system performance
RAID Technology	Highly flexible disk subsystem that offers several levels of redundancy and performance. Uses a specialized software and/or hardware RAID controller. *Advantages* • Increased performance and additive capacity of spanning (striping) • Redundancy and performance advantages of mirroring/duplexing • Failure of single component often undetectable with respect to overall system performance • Increased efficiency with advanced RAID controllers (can provide redundancy with fewer drives than mirroring/duplexing) *Disadvantages* • Advanced calculations in higher RAID levels carry high computational overhead, which can affect I/O applications

Figure 3.6-14. Disk configurations-advantages and disadvantages.

There is also a slight performance advantage associated with using separate logical drives and manually distributing your data files between them. Audio inventory can often be distributed in such a way that normal system use does not hit just a single drive but creates balanced activity between all the drives so more heads are utilized simultaneously.

It is possible to combine mirroring/duplexing with spanning/striping. In this configuration, you span a set of drives, each of which has its own associated mirror drive. This gives you the increased performance and the additive capacity of spanning/striping as well as the redundancy and performance advantages of mirroring/duplexing. In fact, this arrangement performs so well that it is possible for such a system to experience multiple drive failures without significantly affecting system performance. This makes it very important to diligently monitor the status of your drive sub-system and correct any failures as soon as possible. This configuration offers the best of both the performance and redundancy worlds, but with a one drive for the price of two price tag.

Figure 3.6-14 summarizes the relative merits of various disk storage configurations.

Raid

Another disk arrangement, known as *Redundant Array of Inexpensive Disks* (RAID), has been around for a number of years but has recently become very popular. These disk sub-systems are usually comprised of a specialized multi-channel disk controller combined with three or more disk drives. The reason for their popularity is that they are capable of providing high capacity, excellent performance and data redundancy with fewer disks than other configurations mentioned previously.

RAID units can function in several modes referred to as RAID Levels. The various levels offer different combinations of capacity, performance and redundancy. *RAID Level 0* also known as disk striping, provides scalable capacity and high performance but offers no redundancy. *RAID Level 1* also known as disk mirroring, provides redundancy but doubles the number of disks required for a given capacity. RAID controllers can also provide a combination of RAID 0 + 1, which provides scalable capacity, excellent performance and redundancy, but still gives away one half of the capacity to accomplish the redundancy. *RAID Level 2* uses multiple check disks to allow error correction without having to completely duplicate all data. This can provide fully redundant storage with less than double the number of disks, but since higher RAID levels are even more efficient, Level 2 is rarely used. *RAID Level 3* provides redundancy by using nine synchronized drives with a bit of each byte written to the first eight along with a parity bit that gets written to a ninth dedicated parity drive. Level 3 can provide excellent performance, but storing parity at the byte level requires a large number of synchronized drives, which add considerable expense of the system.

The real efficiency comes with *RAID Levels 4* and *5*. In both of these modes, the data stream is sub-divided into smaller blocks (which can be many bytes) with each block being written to a separate disk just as in disk striping (Level 0). But in addition, the blocks written to each drive are compared by an algorithm that generates parity data that is written to another drive. If any drive in the stripe fails, this parity information is combined with the other good drives to recreate the data from the missing drive. When the failed drive is replaced, a similar process is used to rebuild the new disk to the original data structure.

The main difference between the two levels is that Level 4 always stores the parity information for every stripe on a single drive, while Level 5 stores parity information for each stripe on a different drive. RAID Level 4 can only access one stripe at a time limiting its effective I/O for random transfers; however, its parity calculations are somewhat simpler, making it effective for long sustained sequential operations as in single user multimedia applications. RAID Level 5 is able to process multiple stripes simultaneously, making it better suited for random scattered disk requests as in multi user centralized storage environments. The parity structure of RAID Level 5 is a bit more complicated than other levels, but given the proper horsepower and caching to offset this overhead, it is the best choice for a wide variety of file I/O operations. Figure 3.6-15 summarizes the features of the various RAID Levels.

With all of the administration associated with striping along with parity calculations and tracking, these higher RAID levels carry a significant computational overhead. Where these computations take place and how these functions are implemented has a direct bearing on a specific system's suitability for demanding I/O applications.

RAID systems are available in three basic implementations—software RAID, independent hardware RAID and disk controller based RAID.

A software RAID does not use any special hardware but instead utilizes special software running on the server's CPU to manage the striping and parity requirements of the disk array. Simple disk spanning and mirroring functions such as those integrated into the NetWare and Windows NT operating systems are quite effective since they don't add much of a load to the host. However, with higher RAID levels like 4 and 5, this overhead can be quite hefty and unless the server operating system employs some scalable resources to handle the extra load (like the ability to use multiple processors), this implementation will most often consume a significant portion of the server's horsepower and ultimately degrade the overall system performance.

An independent hardware RAID is a self contained unit external to the host PC that communicates with the server through a standard disk controller as if it were a single large disk drive. These external units are very powerful and usually include their own Pentium or RISC processor to handle the RAID management. They can be loaded with large amounts of RAM for drives wear so badly that they would rumble loudly then come to a horrible grinding halt.

While this sort of problem has largely disappeared through improved design and manufacturing processes, mechanical problems associated with handling remain.

Disk drives are extremely sensitive to shock and vibration. A short fall to a workbench or floor is often enough to destroy a drive. The manufacturers have addressed these kinds of problems through improved packaging and handling procedures, but what shipping companies and stock room personnel do with the drive before it gets to you is another story. The best advice for purchasing disk drives is *caveat emptor*—buyers beware—there is probably a good reason for those super deals.

Once the drive has made it to the final destination and is up and running, the biggest environmental factor is heat. Disk drive cabinets must have adequate cooling fans and the fan filters must be kept clean. Air flow through the enclosure should be free flowing and unimpeded by doors or covers. Some cabinets offer temperature sensitive fans that will increase flow as the temperature rises. Others provide temperature monitoring devices and alarms to alert maintenance personnel of a problem. Even though most drives are rated to operate in temperatures up to 50°C (122°F), the cooler they are, the longer they will run.

In an expected lifetime of 100 years, failures within 6 months to a year could be considered infant mortality—certainly within the hardware warranty. But a radio or TV station can become very dependent on such a system in that time frame and will trust it with a considerable amount of irreplaceable data. It is therefore imperative that anyone planning to invest in a station automation system be aware of these issues and be prepared for such an occurrence.

There are several ways to protect data from drive failure. The most basic approach is to use a tape backup device. These are available in a number of different formats, the most popular currently being the 4 mm DAT drives.

Backup software is used to stream the hard disk data onto the tape, which can then be removed and stored safely in another location until needed. The biggest problem with these devices is that the data transfer rate is very slow compared to a disk drive, so backup sessions can take long periods of time. If the system is a single stand-alone computer, you might be required to exit the automation application to run the backup, producing considerable down time.

The data stored on tape is a snapshot of the files at the time of the backup, so it is only as current as the last backup. While many backup software packages allow you to schedule periodic backup sessions, they will usually lag behind the rapid changes made daily in a broadcast automation system.

Also, because the data on the tape must be accessed linearly and sequentially, restoring even a single file can be very time consuming. With these facts in mind, tape based backups should be not be considered "on-line" backup but only for recovering from catastrophic failure.

A much more transparent way to achieve disk redundancy is through *disk mirroring*. With this configuration, the backup device is an identical disk drive. The data on the first drive is automatically replicated onto the second drive, either by the operating system (as in the case of Novell NetWare) or via a special disk controller (as in RAID technology). Either disk can fail and the operation will be unaffected because the system will just use the other drive. It is also possible to mirror more than two disks for additional redundancy; however, this is usually not very cost effective.

Disk mirroring can also enhance system performance during disk reads. Because two (or more) copies of the data are available on independent drives, the controller can pull from both, using whichever drive is better positioned to service a particular request. This increases the number of heads available to gather data. The downside of disk mirroring is that the disk requirements for a given capacity are doubled. For systems requiring large capacity, this can be very costly.

An extension of disk mirroring is known as *disk duplexing*. This is the same as mirroring except that each mirror drive also has its own controller and cable. Disk duplexing adds another level of hardware fault tolerance, allowing the system to be unaffected by a failure of any part of the disk sub-system including the drive, controller, or cable.

Disk duplexing achieves a slightly higher read performance benefit over disk mirroring, since in addition to the multiple heads available to gather data, there are now multiple data buses to carry the data to the host. In addition, this configuration will also provide a write performance boost over mirroring, since the multiple controllers are able to cache and write data to both drives simultaneously.

To increase disk performance even further, a disk configuration called *spanning* or *striping* is often employed. The idea behind this scheme is to sub-divide the data between two or more drives. The operating system or a specialized disk controller takes the data stream and divides it into smaller units then writes one unit to each drive. Because each drive is able to write (or read) simultaneously with the other drives, the disk I/O throughput is increased by a factor equal to the number of drives. This also means the total capacity is equal to the sum of all the drives.

However, there is no redundancy with this configuration. It actually puts a facility at greater risk of overall failure. If any of the drives in the span fail, all of the data on all of the drives is lost. Since more drives increases the total parts count, the probability of failure due to one of the components increases.

In a situation with only a tape backup device or possibly no backup at all, it would be wise to configure each physical disk as a separate logical drive. That way, if one disk drive fails, only the contents of that drive are lost. And, if restoring from a tape drive a single disk's contents will take far less time to restore than the entire disk collection.

	Sampling Rate kHz	KBit Rate	Compression Ratio	MB per sterec minute Divide by 2 for Mono	Stereo hours per GB Multiply by 2 for Stereo
Linear	32.0	N/A	N/A	7.32	2.33
	44.1	N/A	N/A	10.09	1.69
	48.0	N/A	N/A	10.99	1.55
Dolby AC-2	32.0	N/A	6.0	1.22	13.98
	44.1	N/A	6.0	1.68	10.15
	48.0	N/A	6.0	1.83	9.32
ISO/MPEG Layer 1	32.0	32	16.0	0.46	37.28
	32.0	64	8.0	0.92	18.64
	32.0	96	5.3	1.38	12.35
	44.1	32	22.1	0.46	37.37
	44.1	64	11.0	0.92	18.60
	44.1	96	7.4	1.36	12.51
	44.1	128	5.5	1.84	9.30
	48.0	32	24.0	0.46	37.28
	48.0	64	12.0	0.92	18.64
	48.0	96	8.0	1.37	12.43
	48.0	128	6.0	1.83	9.32
	48.0	160	4.8	2.29	7.46
ISO/MPEG Layer 2	32.0	32	16.0	0.46	37.28
	32.0	48	10.7	0.68	24.93
	32.0	56	9.1	0.80	21.20
	32.0	64	8.0	0.92	18.64
	32.0	80	6.4	1.14	14.91
	32.0	96	5.3	1.38	12.35
	44.1	32	22.1	0.46	37.37
	44.1	48	14.7	0.69	24.86
	44.1	56	12.6	0.80	21.30
	44.1	64	11.0	0.92	18.60
	44.1	80	8.8	1.15	14.88
	44.1	96	7.4	1.36	12.51
	44.1	112	6.3	1.60	10.65
	44.1	128	5.5	1.84	9.30
	48.0	32	24.0	0.46	37.28
	48.0	48	16	0.69	24.86
	48.0	56	13.7	0.80	21.28
	48.0	64	12.0	0.92	18.64
	48.0	80	9.6	1.14	14.91
	48.0	96	8.0	1.37	12.43
	48.0	112	6.9	1.59	10.72
	48.0	128	6.0	1.83	9.32
	48.0	160	4.8	2.29	7.46

Figure 3.6-13. Disk storage capacity requirements.

$$\left[2\ \text{CHANNELS} \times \frac{\#\ \text{SAMPLES}}{\text{CHANNEL*SEC}} \times \frac{16\ \text{BITS}}{\text{SAMPLE}} \times \frac{1\ \text{BYTE}}{8\ \text{BITS}} \times \frac{1\text{GB}}{1024^3\ \text{BYTES}} \times \frac{3600\ \text{SEC}}{\text{HOUR}} \times \frac{1}{\text{COMPRESSION}} \right]^{-1}$$

$$= \frac{\text{COMPRESSION}}{\text{SAMPLE RATE}} \times 7.456540444 \times 10^4\ \text{HR/GB}$$

Stereo hours per gigabyte

$$2\ \text{CHANNELS} \times \frac{\#\ \text{SAMPLES}}{\text{CHANNEL*SEC}} \times \frac{16\ \text{BITS}}{\text{SAMPLE}} \times \frac{1\ \text{BYTE}}{8\ \text{BITS}} \times \frac{1\ \text{MB}}{1024^2\ \text{BYTES}} \times \frac{60\ \text{SEC}}{\text{MIN}} \times \frac{1}{\text{COMPRESSION}}$$

$$= \frac{\text{SAMPLE RATE}}{\text{COMPRESSION}} \times 2.28881835 \times 10^{-4}\ \text{MB/min}$$

Megabytes per stereo minute

Sample rates are not abbreviated: use 32000, 44100, or 48000.
Use compression=1 for linear audio files.
Dolby compression is always 6.
MPEG compression is variable; consult the "Approximate Storage Requirements" chart from Figure 3.6.13 for the compression ratio that corresponds to the bit rate used.

Figure 3.6-12. Digital audio data rate calculation.

resolution or accuracy of the sample. This determines the quantization (digital) noise floor which relates directly to the overall dynamic range, signal to noise (signal to error) ratio and distortion components of the digitized signal. Sixteen bit quantization provides a theoretical signal to error ratio of 98 dB, which is very near the limit of human perception. The mono/stereo issue is simple—stereo produces twice as much data as mono.

Putting these numbers together, we arrive at the two equations shown in Figure 3.6-12. The first allows us to calculate the number of hours of stereo audio that can be stored per gigabyte of disk capacity. The second gives us the number of megabytes required to store a minute of stereo audio. As you can calculate, digital audio requires a considerable amount of hard disk capacity.

Which brings us to the issue of data compression. The equations of Figure 3.6-12 include a possible compression factor which will reduce the overall disk capacity requirements. The data compression issue is quite complex and is beyond the scope of this discussion. It is, however, discussed in detail in Section 3.7. For the purpose of this discussion, we will indicate that most networked digital audio automation systems use data compression to keep network bandwidth and hard disk storage requirements down. Among the several good compression algorithms available, MPEG and Dolby AC-2 are two of the most popular. Figure 3.6-13 compares Linear PCM, Dolby AC-2 and MPEG compressed digital audio in terms of effective compression ratios and required storage capacity for the sample rates and data rates commonly used by automation systems.

Disk Drive Systems

Now that we have a handle on the capacity we require, we must decide where to store it and what components to use to store it. As described earlier, we can choose to provide hard disk storage in each workstation, in a centralized server, or in both. For reasons stated earlier, a network employing a centralized dedicated file server provides the most complete set of features for a multi-station automation system. We will therefore focus on the disk requirements of such a system. In particular, we will mention some hard disk performance issues then describe several disk configurations, including *disk spanning*, *disk mirroring*, *disk duplexing* and *RAID Technology*, and outline their advantages, shortcomings and levels of redundancy.

The greatest point of failure in any computer based system is the hard disk drive. The issue is not *if* a disk drive will fail, but *when* it will fail. Current premium quality high performance drives are quoting figures of 1 million hours *mean time between failure* (MTBF), which is over 100 years. However, a large percentage of the high capacity drives being used for multimedia systems are falling far short of this mark—many failing within the first 6 months of operation. There appear to be several reasons for this. A few stem from manufacturing problems while most are related to environmental conditions such as heat, shock and vibration.

Although the drive manufacturers are continuously improving the reliability of disk drives by reducing the internal parts count and using higher levels of integration in the on board electronics, the ultimate storage mechanism is mechanical, and therefore subject to mechanical failure.

In the early releases of high capacity drives, primarily the 4 and 9 GB sizes, several manufacturers had problems with the bearings used to support the spindle and its associated media platters. The bearings would wear prematurely causing the spindle and platters to shift and wobble, thereby degrading track alignment and causing loss of data. It was not uncommon to see for two nodes to end up transmitting at the same time. Packets take a finite amount of time to move down the wire, even when traveling near the speed of light. It is therefore possible that a workstation on one end of the wire might start transmitting and before the signal reaches a workstation at the other end of the wire, this workstation checks the network and believing it is free, also begins transmitting. When these packets collide, the communication becomes garbled and useless. This is what *collision detection* is designed to detect and correct. During transmission, a node continues to listen. If at any time during a transmission a node detects another node's signal, it stops transmitting and sends a jamming packet to alert all workstations that a *collision* has occurred. Each transmitting node then waits for a short random period then tries the transmission again. If there is another collision, the node waits an even longer period then tries again. This continues until the back-off-time reaches some maximum limit. As crude as the process sounds, this contention access method is very effective, due to the available bandwidth of the network and the small packet size of Ethernet (1518 bytes).

CSMA/CD is not deterministic. Access to the network is not guaranteed, only statistically assured. For multimedia networks, this might not be good enough. As traffic increases on an Ethernet network, so do collisions. Collisions cause extra *latency* or time delay between the request for data and the receipt of data. Real time processes like digital audio and video require that data packets arrive at very regular intervals and thus are very sensitive to latency. Even with large amounts of available bandwidth (such as with 100BaseTX), it is possible that a workstation requesting audio data from the server might encounter enough sustained collisions on a busy network to delay the packets to the point of causing dropouts in the audio playback. If loaded even further, the number of collisions can increase to the point that packets cannot successfully be exchanged and are lost, resulting in major problems.

Token passing schemes as used in Token Ring can provide determinism since every node is guaranteed regular access to the network. But as described earlier, a node must have possession of the token in order to transmit, and is then only allowed to send one packet before passing the token to the next node. To send another packet, the workstation must wait for the token to be passed all the way around the network even if the other workstations have nothing to transmit. If the network is large, the latency associated with this mechanism might also be unacceptable.

This is where 100BaseVG-AnyLAN excels. 100VG uses a new access control mechanism known as demand priority protocol (DPP) which places the hub in the role of traffic cop. When a workstation has data for transmission, it registers a request with the hub that places the request in a queue. The hub grants access to the network according to the queue and thus provides guaranteed access in an orderly, first come first served basis. And the latency associated with this mechanism is low because no time is wasted passing control through every node on the network whether they need access or not. If only one workstation is requesting the right to transmit, the hub will continue to give it clearance to transmit as long as no other requests are received. This provides a much more efficient use of the network since only workstations that need to move packets use any resources.

Another important feature of DPP is the ability to prioritize network access. One hundred VG hubs support connections at two separate priority levels, normal and high. Each has its own associated queue. Requests in the high priority queue are acknowledged first, providing minimum latency to connections running time sensitive applications. Requests from the normal queue. can be interrupted by high priority requests, but are ultimately protected by an internal mechanism that guarantees they will receive timely servicing even in the presence of large numbers of high priority requests. This feature allows a single 100 VG segment to service time sensitive applications like automation workstations as well as less critical applications like scheduling and billing systems without the fear of affecting the audio performance.

CENTRALIZED STORAGE REQUIREMENTS

Now that we have a better idea of what a station automation system is and how individual units can be networked to enhance efficiency, we should turn our attention to the storage requirements of a networked digital audio automation system.

Analog audio is encoded for digital storage through a modulation technique known as *sampling*. This is sort of like taking a snapshot of the audio waveform at regular intervals. Sampling a continuous waveform produces a stream of pulses, and is thus a form of pulse modulation. The amplitude of each pulse is then converted into a digital number that can be stored as data, which is known as *coding*. Therefore, the overall process of sampling an analog signal into a stream of digital data is known as *pulse coded modulation* (PCM) and is often referred to as *linear PCM* to emphasize that no data compression or reduction schemes have been applied. The amount of data generated by this process depends on how fast we sample, how many bits are devoted to each sample and how many signals we sample simultaneously (mono or stereo).

In order to completely describe the incoming audio waveform, the *Nyquist sampling theory* states that the sampling must occur at a rate at least twice as fast as the highest frequency in the incoming audio. Due to the limitations of real world filtering required to keep samples from being contaminated (known as anti-aliasing filters), the sample rate must actually be slightly higher than twice the highest input frequency expected. Thirty-two kHz sampling has a reasonable frequency response up to 15 kHz, 44.1 kHz sampling is good to about 20 kHz and 48 kHz sampling can produce frequencies up to 22 kHz or so.

The number of bits used to describe each sample is known as the level of quantization and determines the

100Mbps networks	Description, advantages, and disadvantages
100BaseTX	Fast Ethernet uses Carrier Sense Multiple Access with Collision Detection (CSMA/CD) contention-based media access control. *Advantages* • Helps bandwidth issue (but performance in multimedia applications is not bandwidth issue–real-time processes are more affected by latency) *Disadvantages* • Network slows down when loaded • Packets can be lost from "aborts due to excessive collisions" • Hub-to-node connections are limited to 100 meters (328 feet) with CAT5 wiring • Hub-to-hub connections are limited to five meters (16 feet) • Total network span is limited to 250 meters
100BaseVG-AnyLAN	Best of Ethernet combined with the determinism of Token Ring. No collision mechanism; uses centrally controlled round-robin polling network access method called Demand Priority Protocol (DPP). Each node's request is queued at the hub. The hub responds to the requests in order of priority and time received. When access is granted, node gets use of wire's full bandwidth. *Advantages* • Perfect for feeding the sustained data streams encountered in multimedia applications • Loading has minimal effect on interface speed • Supports high/low priority so that workstations can be assigned priority • Hub-to-node connections can be up to 150 meters (492 feet) with CAT5 wiring • Hub-to-hub connections can also be up to 150 meters (492 feet) • Three hub-to-hub uplinks are possible–network diameter of over 2000 feet possible with standard equipment *Disadvantages* • Not as popular as 100BaseTX–fewer suppliers

Figure 3.6-11. Comparison of 100 megabit network technologies.

particularly for multimedia networks such as station automation systems. This network can support both Ethernet and Token Ring frame types and cabling schemes and is therefore a logical upgrade from either environment. It is, however, different enough from both that it is covered by its own standard, IEEE 802.12. 100VG (as it is often abbreviated) is a physical star network that utilizes 4 pair of voice grade wires or better to connect a workstation or server to the hub. This allows a simple upgrade to the 100 Mbps world for nearly any existing installation. Using voice grade or Category 3 cable, this connection can be up to 100 m. If using Category 5 wiring, this connection can be extended to 150 m. Also, up to 3 hubs can be interconnected with the hub-to-hub distance following the same rules. This provides for networks to span as much as 750 m without any special equipment. The most interesting feature of this network is the use of intelligent hubs that manage how the workstations are able to access the network. To understand the benefits of this new media access control mechanism, we must first see how Ethernet manages transmissions.

No matter what the physical topology or speed, all IEEE 802.3 Ethernet networks operate as bus networks. This is obvious with the coax based networks, but is also true of the star wired twisted pair variations. These networks employ hubs that give all the benefits of point-to-point node-to-hub connections but internally provide a shared bus for network communications. And as discussed earlier, for a bus network to function, it must include a control mechanism to ensure only a single transmitter accesses the network at a time. In Ethernet, this mechanism is known as *carrier sense multiple access with collision detection* or CSMA/CD. What this means is that multiple workstations have equal access to the same media but must first sense that the network is free (has no carrier) before they begin to transmit. But it is still possible

Network Topology	Ethernet Class Supported	Description, Advantages, and Disadvantages
Bus	10Base2 10Base5	Simple workstation/server connections; best for small installations. *Advantages* • No hub is required • Good RF shielding (can be improved with grounding and ferrite cores) *Disadvantages* • Sensitive to proper termination • Single point failure can kill entire segment • Load balancing across multiple segments can get messy
Star	10BaseT 100BaseTX 100BaseVG-AnyLAN	Requires a hub/router per segment, so best for medium-large installations. *Advantages* • Single point failure only kills single client • Easy to cable and terminate if quality cable and connectors are used • UTP cable has reasonable RF/noise immunity due to twisted pair, can be improved with STP • Easy to load balance by repatching at hubs *Disadvantages* • Hubs can be costly for small installations

Figure 3.6-10. Features of common network topologies.

from 10 Mbps based on a 200 m (approximate) segment. The specification defines the maximum supported length as 185 m with a maximum of 30 nodes spaced a minimum of 0.5 m apart. Workstations are attached directly to the backbone by a BNC "T" connector placed on the BNC port of their NIC and then inserted into the backbone. To make these connections, the backbone must be routed past the back of every networked PC. Due to the number of connectors involved, the potential for cable failure is sizeable. In a bus network, a single point failure will kill the entire network. However, the advantage of the less expensive, more flexible and easier to work with RG-58 style cable can often outweigh this concern and makes Thinnet a viable cabling scheme for small to moderate sized networks.

10BaseT, stands for 10 Mbps based on twisted pair wiring. 10BaseT is physically a star network which utilizes UTP to make point-to-point connections between each workstation or server and a centrally located hub. The UTP cable, typically used, includes 4 separate pairs in a single sheath and is rated by category. 10BaseT requires UTP cable rated as *Category 3* or higher. 10BaseT connections require only 2 of the pairs (one transmit, one receive), and use standard easy to apply RJ-45 connectors. Cable runs are limited to 100 m, making the effective distance from workstation to hub to server up to 200 m. The use of active hubs provides isolation of each connection, reducing the sensitivity to single point failure, and provides convenient patch points, which can ease expansion and load balancing. The low cost of the cabling and the ease of installing, connecting and maintaining these networks has made 10BaseT the most popular cabling arrangement for mid to large size networks. Figure 3.6-10 presents a quick comparison of these common 10 Mbps Ethernet configurations.

In addition to the 10 Mbps offerings, Ethernet has recently advanced to the 100 Mbps arena. 100BaseT, dubbed *Fast Ethernet*, includes three different cabling standards: 100BaseFX, 100BaseTX, and 100BaseT4 (see Figure 3.6-11). All use a star physical topology, with each node having a point-to-point connection with the hub. With 100BaseFX, these connections are made with two strands of fiber-optic cable. 100BaseTX uses two twisted pairs of Category 5 UTP or STP. 100BaseT4 uses four twisted pairs of Category 3, 4 or 5. For new installations where no existing cable needs to be integrated, either the FX and TX options make good sense since all new cabling must be installed anyway. For existing 10BaseT installations that utilize Category 3 wiring, 100BaseT4 allows a simple upgrade path to 100 Mb without having to replace the wiring. In practice, 100BaseTX using Category 5 wiring has become the standard of choice and there has been relatively little application of the others. 100BaseTX links are limited to 100 m with an overall segment length limit of 250 m.

Another 100 Mbps standard called 100BaseVG-AnyLAN has emerged with some excellent features, point to point transfers such as workstation to server communications.

A star network offers some distinct advantages over other topologies. Point-to-point communications between the workstation (or server) and hub can provide more efficient full duplex communications. The connection ports on a hub provide a central patch point for easy expansion and reorganization of the network. Most all hubs in use today are active devices that include signal conditioning electronics to clean and reshape waveforms, which improves the performance, reliability and range of the network. The conditioning electronics also buffer the ports making the network less sensitive to a single point cable failure. The hub can just ignore the disconnected (or shorted) port without affecting the rest of the connections. And most hubs also include traffic monitoring and management functions, which allow system administrators to analyze and *fine tune* their networks. The one disadvantage of this topology is the cost of the hub itself, which for small networks may not be warranted.

Network Cabling and Hardware

So what does all this mean to the station engineer who needs to install and connect the station automation system? To get to the real world nuts and bolts of networking, we must discuss cabling standards.

There are three popular local area network (LAN) cabling standards: Ethernet, Token Ring and ARCnet. Two of these, Ethernet and Token Ring, are defined as international standards by the Institute of Electrical and Electronics Engineers (IEEE). ARCnet is not an IEEE standard but has a substantial installed base and therefore deserves some mention.

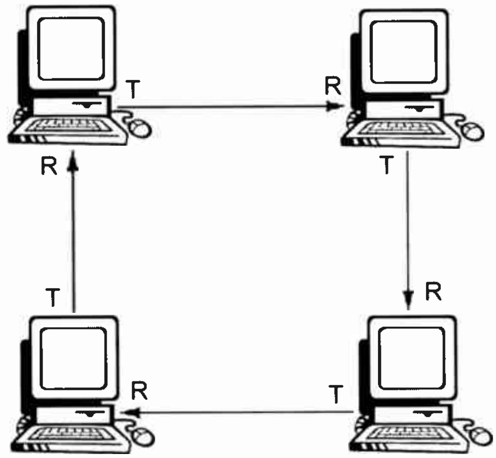
ARCnet utilizes a star physical topology with 93 Ω coaxial cable (RG-62 style) connecting workstations and servers to a mixture of passive and active hubs. The use of passive hubs makes the network inexpensive and simple to wire, however, the network logically functions as a bus and is therefore sensitive to proper termination and cable failure. Active hubs isolate connections to protect the network from single point cable failures and extend the span of the network (up to 20,000 ft. overall). This network was the first supported by Novell NetWare and still retains a respectable presence, mostly from early NetWare networks. ARCnet originally provided a data rate of only 2.5 Mbps as compared to Ethernet's 10 Mbps, but at lower cost, making it a reasonable trade off for many installations. It has since been upgraded to provide 20 Mbps, but ARCnet has never been able to overcome the growing popularity of Ethernet and is rarely seen in new LAN installations.

Token Ring networks can provide data rates of either 4 Mbps or 16 Mbps over either *shielded twisted pair* (STP) or more recently *unshielded twisted pair* (UTP) cabling. The network is physically wired in a star topology with hubs, also know as *multi-station access units* (MSAU), providing a logical ring connection between *nodes*. Access to transmit on the ring is controlled by a special data frame called a *token*, which is circulated around the ring from the transmitter of one node to the receiver of the next and so on. If a workstation has data to transmit, it must wait until it receives possession of the token. When the workstation has transmitted its data frame, it must then release the token to the next workstation in the ring. This is a very effective access control mechanism because it is deterministic, meaning every node is given equal opportunity to access the network on a regular basis. Token access control is, however, a complicated mechanism to implement because it must include provisions for detecting and dealing with lost tokens and undelivered packets that could circulate the ring for ever. IBM is the originator of Token Ring and still holds the patents for the most common chipset used for Token Ring NICs. This has caused Token Ring to be viewed as more proprietary than other technologies. And while Token Ring has enjoyed widespread popularity particularly in large corporate and educational environments, it is more expensive than Ethernet for a similar level of performance, and is therefore less likely to be used in new systems. Token Ring is defined in the IEEE 802.5 Standard.

Ethernet was the first and is by far the most favored network cabling standard in use. It has a number of different physical implementations including 10Base5, 10Base2 and 10BaseT which are 10 Mbps technologies, and 100BaseTX, 100BaseT4 and 100BaseFX which are 100 Mbps technologies. Another standard known as 100BaseVG-AnyLAN is a special 100 Mbps technology that is somewhat of a cross between Ethernet and Token Ring and offers some very interesting features.

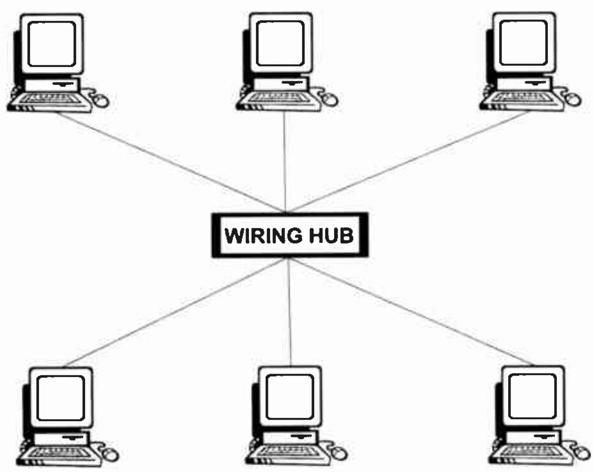
10Base5 is physically a bus topology that utilizes a 0.5 inch diameter 50 Ω coaxial cable (RG-8 style) as a trunk or backbone. The size of this cable has led to 10Base5 being dubbed *thick wire Ethernet* or *thicknet*. The coaxial backbone acts as a transmission line, and therefore must be terminated in its characteristic impedance (a 50 Ω load) at each end to minimize the reflection of signals that may be traveling on it. Workstations are attached to the network through NICs with *attachment unit interface* (AUI) ports that connect through 15 conductor cable to transceivers called *media attachment units* (MAU) that are either inserted into or tapped onto the trunk cable. The trunk can be up to 500 m long, hence the IEEE designation 10Base5—meaning 10 Mbps based on a 500 m segment. A single segment can support up to 100 nodes which must be spaced a minimum of 2.5 m apart. The connectors, cable and attachment units required for 10Base5 are expensive and difficult to work with while offering few advantages over other connection schemes. Thicknet is therefore rarely seen in new installations.

10Base2 is also physically a bus network that uses a 0.2 in. 50 Ω coaxial cable (RG-58 style) as the main trunk. Due to the thinner cable, this configuration is often called *thin wire Ethernet* or *thinnet*, and due to the lower cost of the cable, it is often referred to as *cheapernet*. The 10Base2 IEEE designation comes



T=TRANSMIT R=RECEIVE

(d)



(e)

Figure 3.6-8. Physical topologies: d) ring, e) star.

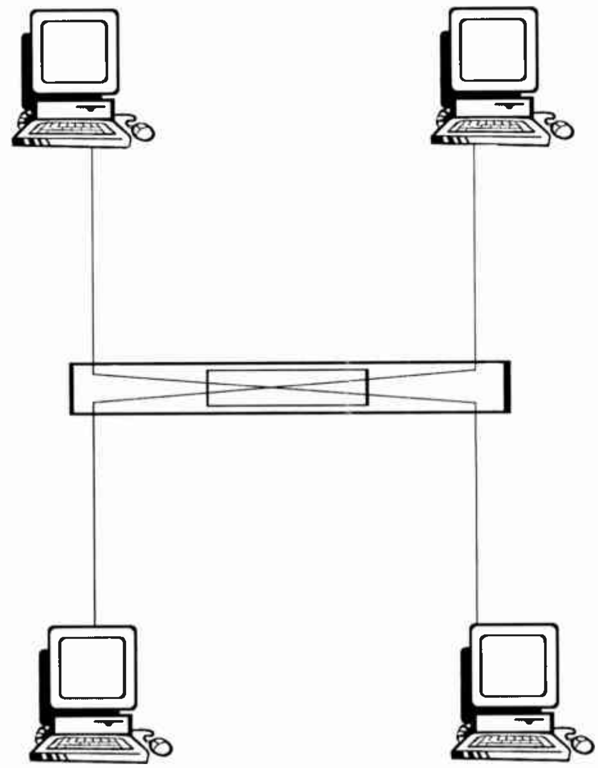
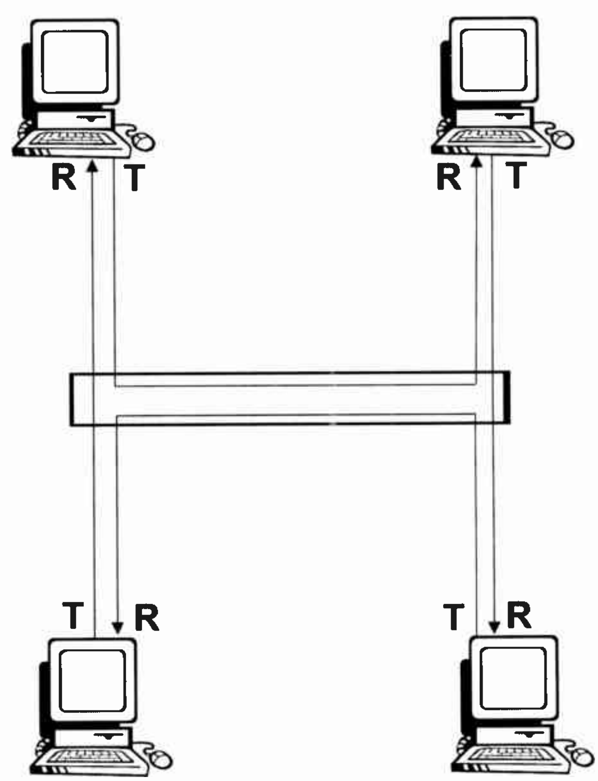


Figure 3.6-9. Logical topologies: a) star as bus, b) star as ring.



(b)

can be situations where a packet must travel the entire circumference of the ring to reach the workstation immediately adjacent to it. If the network is very large, this can add a considerable delay to the exchange of data, resulting in a sluggish network.

By far, the most popular implementation of the point-to-point link network is the *star* topology (see Figure 3.6-8 (e)). In this arrangement, all workstations/ servers have a point-to-point connection with a centrally located wiring cabinet called a hub, which provides the signal routing between the nodes. How the hub internally connects the nodes determines the *logical topology* of the network, or, how the network works. While the network might physically be a star topology using point-to-point connections, the hub might logically connect the workstations as a ring or bus as shown in Figures 3.6-9 (a) and (b). Advanced switching hubs can provide a form of mesh logical topology which can greatly increase the efficiency of

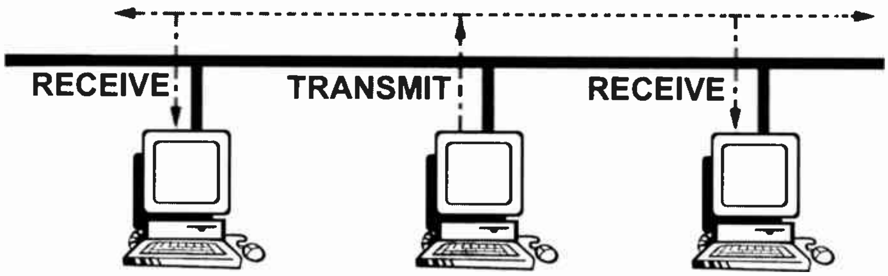
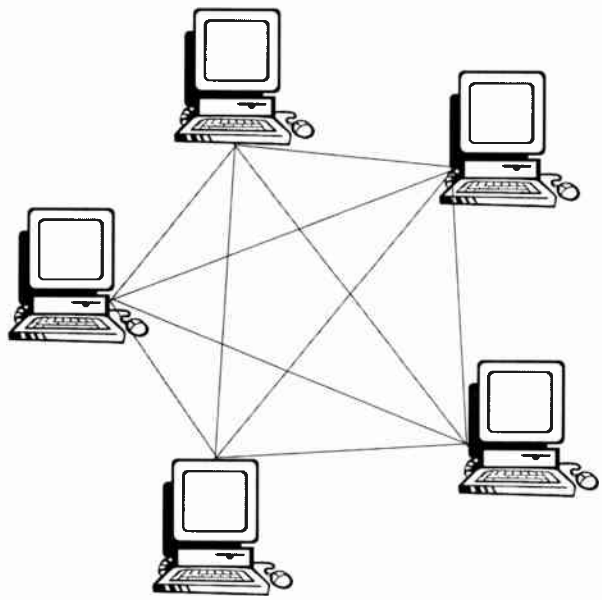


Figure 3.6-8. Physical topologies: a) bus, b) point to point, c) mesh.

(b)



(c)

Each of these network architectures provides useful connectivity between workstations and has its place in station automation. But, all things considered, the dedicated server based network offers the best collection of features from which to build a multi-station automation system.

Network Topologies

With a basic understanding of some of the operational benefits that PC networking can bring to our overall station automation picture, it makes sense to discuss how this networking is physically accomplished. The way cables are arranged to provide interconnection in a network is known as its *shape* or *physical topology*. The way the signals are actually routed throughout the network is known as its *logical topology*. The physical topology and logical topology may not necessarily be the same, as we will see shortly.

There are two building blocks for all physical topologies: the *bus* and the *point-to-point* link.

The bus topology uses a single cable that each workstation and server attaches to and shares. When any *node* (connected devices are often referred to nodes) transmits, it broadcasts over this transmission line with every node hearing every transmission. Communications would become hopelessly confused if multiple workstations tried to transmit on this shared line simultaneously; therefore, bus networks must employ a control mechanism to make sure only one transmitter is active at a time. The bus topology can be easily expanded by making the bus cable longer and attaching more workstations up to the electrical limitations of the cable. It is also possible to expand beyond this limit through the use of repeater electronics and multiple cable segments. Figure 3.6-8 (a) represents the basic concept of the bus topology.

Point-to-point networks, on the other hand, do not broadcast on a shared cable, but as the name implies, transmit to exactly one receiver and receive from exactly one transmitter. As Figure 3.6-8 (b) illustrates, the transmitter of one workstation is connected directly to the receiver of another with the transmitter of that workstation connected back to the receiver of the first. Because transmission and reception occur over separate wires, this topology has the advantage of supporting *duplex operation* (simultaneous transmit and receive), which doubles the communication speed. However, point-to point networks are somewhat more complicated to expand. One way to add workstations is to provide a point-to-point connection between every pair of workstations in the network as shown in Figure 3.6-8 (c). It is obvious that this would become overly complex for even a small group of computers. Another way to expand a point-to-point link network is to connect the workstations in a *ring* arrangement as shown in Figure 3.6-8 (d). The transmitter of one workstation is connected to the receive port of the next workstation in the ring. Data, transmitted in individually addressed units called *packets*, travels from point-to-point-to-point around the ring until it reaches its destination address. With full duplex communications, this can be an effective arrangement, however, since packets are transferred around the ring in a single direction, there tion can suffer during a transfer. In a peer-to-peer network, each unit typically runs from its own database, requiring considerable file maintenance to keep the multiple separate inventories organized.

An audio server based network usually consists of one or more workstations communicating with a dedicated and often proprietary computer that houses both the audio processing hardware (DSP boards) and audio file hard disk storage (see Figure 3.6-6). The workstations act as control interfaces that command the central audio server to perform all record and play functions. A nice feature of the arrangement is that all the audio data is stored centrally, making it instantly accessible by any workstation. Audio recorded by a production workstation can be immediately played by an on-air workstation without requiring any file copying to the local hard drive. A serious disadvantage to this network configuration is that all audio processing is performed by the audio server, which has a physical limit to the number of audio channels it can contain (typically 8 or less). This limits the number of workstations that can be supported by the system and makes it difficult and expensive to expand. Also, because all audio inputs and outputs occur at the audio server, audio interconnect wiring is required to and from each studio adding to the installation complexity and potential points of failure. And since the workstations contain no audio hardware themselves, they have no functionality in the event of an audio server or network failure.

A dedicated file server based network combines the good points of these first two architectures. The network generally consists of one or more workstations linked to a central high performance, non-proprietary computer that runs a standard high performance NOS like Novell NetWare, Unix NFS, or Windows NT Server. The server presents itself to the rest of the system as a large shared disk drive and manages the digital storage of all audio and system data. As with the audio server concept, the audio is stored centrally and is therefore instantly accessible by any of the workstations. The major advantage this configuration has over the audio server configuration is that the audio processing hardware (DSP boards) are located in each of the individual workstations as illustrated in Figure 3.6-7. This places the audio inputs and outputs at the workstation with the network used to transfer packets of digitized audio data between the workstation and server. Not only does this allow all workstations to share the centrally stored audio, but it also provides

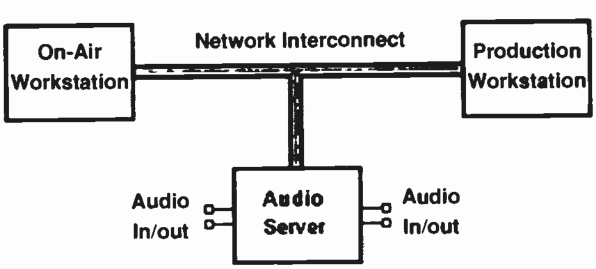


Figure 3.6-6. Audio server network.

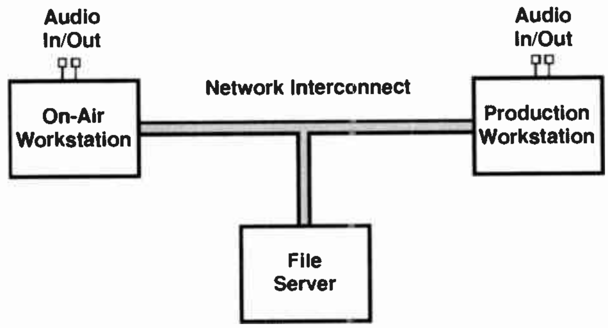


Figure 3.6-7. Dedicated file server network.

audio distribution via the data network. An audio source local to a workstation at one end of a facility can be recorded and then immediately played by a workstation at the other end of the facility. A practical example of this is a workstation installed in an equipment room with the station's satellite receivers and programmed to automatically record multiple news feeds throughout the day. Workstations anywhere within the facility can access, edit, and play portions of these feeds without having any direct audio feed from the receivers themselves. The interesting thing is that the complete interconnect for distributing the normal system data and audio is a single piece of coax or twisted pair cable!

The fact that the file server contains no audio hardware is also very significant to server functionality, because this means it is not bound by any physical number of channels. The server is designed to service large numbers of data streams bound only by performance and capacity of its disk (input/output) subsystem and the speed of the network. With proper server design, any number of workstations/audio channels can be supported, making this type of network scalable.

Another excellent benefit of the dedicated file server based network is the various data redundancy (backup) options available. With a number of workstations relying on the file server to store all of their work, it is important that this storage remains online at all times and is protected in the event of catastrophic hardware failure such as a hard disk failure. NOS software for dedicated file servers is quite sophisticated and generally contains special fault tolerant features to protect the integrity of the data being stored. Features like disk mirroring, server mirroring and Redundant Array of Inexpensive Disks (RAID) are available, along with support for a considerable range of industry standard hardware components designed to provide performance and redundancy. Also, the dedicated file server architecture is able to provide a level of redundancy via the workstation. The primary mode of operation is for each workstation to record and play directly to and from the server; however, because each workstation contains its own audio processing hardware, local hard disk storage can be included for emergency use in the event of file server or network maintenance or failure.

Figure 3.6-4. Touchscreen controlled audio playback buttons integrated with text.

as PC networking. A single stand-alone PC with the features described thus far can do an excellent job providing the basic production, live-assist and automation features required for a single station. If you need to automate multiple stations, you have two choices: try to *squeeze* the additional resources from a single workstation or add additional workstations to handle the additional stations. Using a single workstation simplifies inventory management and allows the production and scheduling to be shared; however, as you try to do more and more within a single box, the unit's capabilities will be spread thinner and thinner until the overall performance is no longer satisfactory. If you add more independent units, you maintain each system's performance level but add the burden of managing multiple inventories, multiple production facilities and multiple scheduling systems. To achieve the best of these two scenarios we would need to interconnect the workstations so they could share audio inventory, share production elements and share scheduling resources while they each provided high performance live-assist and/or automation features to their individual stations. This kind of interconnection is exactly what PC networks are designed to provide.

Networking the individual workstations together enables them to move audio, automation schedules, as-play logs, scripts and wire copy text, configuration and control data and software updates from unit to unit without having to physically transport the data via removable media (floppy disk, Bernoulli cartridge, writable CD, magnetooptical disk or data tape). Eliminating the time and effort associated with such a *sneaker net* is usually by itself enough to justify the expense of the network hardware and software. But in addition to providing a conduit to expedite the exchange of information between one unit and another, networking can actually increase the overall efficiency of a group of workstations by eliminating unnecessary replication of resources and by allowing individual workstations to be specialized for particular functions. For instance, it is usually more efficient in both equipment and manpower, to equip a single workstation with the production facilities to create spots for all of the stations than it is to perform production at each of the on-air workstations. This also makes the on-air workstations more efficient by allowing them to devote their resources to their primary on-air functions. The same is true of other group wide functions such as wire service integration, news feed collection, scheduling and overall inventory management.

Network Architectures

A number of computer network architectures are presently being utilized for broadcast applications. Each has certain advantages and disadvantages. The three most common configurations are the *peer-to-peer* network, the audio (video) server based network, and *dedicated file server* based network.

A peer-to-peer network is simple and inexpensive way to link two or more workstations at the file system level, permitting file transfer between them. This arrangement has achieved reasonable popularity since it only requires the addition of *network interface cards* (NIC) to each workstation and the loading of some relatively simple *network operating system* (NOS) software. The basic concept of a two workstation network is illustrated in Figure 3.6-5. Peer-to-peer networks allow each workstation to access the contents of another workstation's drive; however, the performance of this connection is such that audio files cannot be shared directly. To play or record audio files, the files must reside on that workstation's local hard drive. If an audio file required for playback exists on another workstation's drive, it must be copied across the network to the local hard drive before it can be used. This can add a considerable delay between the production and use of an audio cut if it must be transferred since audio files can be quite large. Also, a workstation's capabilities can be taxed by the overhead required to service the network, and normal on-air opera-

Figure 3.6-5. Peer-to-peer network.

sion of custom serial protocols, be capable of storing and running *control macros* that execute specific command sequences and provide these features through a user programmable mechanism so that each station's unique compliment of equipment can be administered and maintained.

Even if an automation system possesses all of the key features described thus far, it MUST have a clear, well organized, intuitive operator interface or much of its power will go unused. Many of the early automation systems were quite powerful but displayed busy text based screens and required multiple key keyboard commands to operate. The *computer-phobic* shied away from these units, making them useful to only the more computer literate station personnel. Graphic user interfaces (GUI) like Mac O/S, IBM O/S2, Unix X-Windows and various versions of Microsoft's Windows have made it possible to display screens that provide more visual clues as to function and use. The necessary text-based data remains, but it is surrounded with buttons and scroll boxes that allow the operator to navigate the screen and operate the system with a simple pointing device like a mouse or trackball (see Figure 3.6-1).

Some of the more advanced automation systems have taken this a step further by allowing the pointing device to be the operator's finger. This touch screen technology coordinates a touch sensitive x-y matrix placed over the display screen with the graphics displayed through it by the program. The operator can *press* a button on the screen and get the illusion that they are operating a physical button due to the program's resultant visual clues like color change and shadow reversal. This type of interface has engendered PC based automation systems into common use by even the most computer-phobic staff members, and has enhanced the role of the automation system as an on-air live assist device. Screens that present lists of cuts to be played sequentially along with their *hit times*, transition information, a count down clock, a time of day clock and audition functions work well for tightly formatted live assist periods (see Figure 3.6-2). Screens that present panels of programmable playback buttons provide the instantaneous playback capabilities needed for spontaneous live shows (see Figure 3.6-3). And script based programs like newscasts can benefit from systems that allow the integration of text and audio to provide sound bite playback from buttons that are embedded in and scroll with the script (see Figure 3.6-4). Examples such as these illustrate how a well designed human-machine interface can broaden an automation system's usability and enhance its effectiveness.

Figure 3.6-2. Sample of touchscreen controlled live assist graphic user interface.

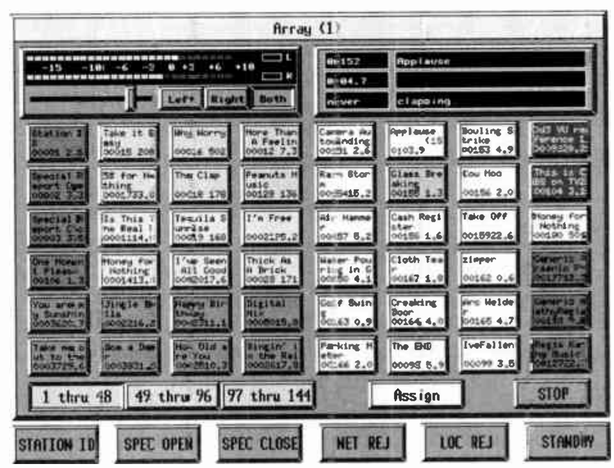


Figure 3.6-3. Sample of touchscreen controlled random access graphic user interface.

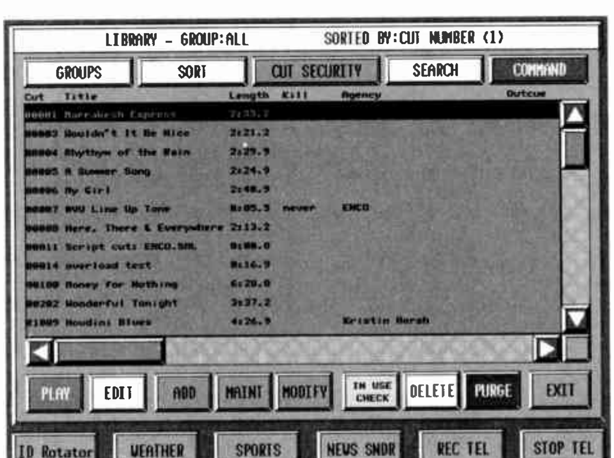


Figure 3.6-1. PC user interface combining audio and database functions.

NETWORKED AUTOMATION SYSTEMS

Of all of the contributions the personal computer has made to advance the state of the art of station automation perhaps none has had as pronounced an effect gramming and thus represent the station's daily program schedule or log. In addition to the actual audio elements, these playlists need to include automation information that instructs the system how to sequence from one to the next automatically. The ability to program lists to cover multiple days or the ability for one list to sequence to another list is also very important.

Because these playlists might typically hold hundreds or possibly thousands of elements, scheduling each and every event and programming the automation manually would be very tedious. Again, the power of the PC comes to the rescue. There are many PC based programs in existence that use database functions combined with user defined sets of rules to manage and schedule commercials/promos (traffic) and music. Each category is handled somewhat differently. Programs that schedule commercials generally take sales department information (orders) and apply rules based on air time rates, time of day, time of year and adjacent programming to generate a daily schedule or log. These programs are usually also tied to the billing department and are often able to accept *as-played* logs from the automation system and reconcile them with the original schedule to generate the appropriate billing invoices. Music scheduling programs on the other hand must choose cuts based on the station's format, time of day, music category and artist, and may apply rules like not allowing two artists from the same music category to play back-to-back or like limiting specific cuts from playing more than once per hour or once per day.

While some automation systems provide these scheduling features as part of an overall comprehensive automation package, it is not necessarily a great advantage to obtain everything from a single vendor. With all the excellent scheduling software that is already available from a multitude of vendors, often the most flexible approach is to choose each package separately based on its individual merits (possibly already familiar or in use). Then select an automation system that can interface with these systems to import and merge the traffic and music schedules and provide seamless integration of scheduling and playback. The automation system then exports an as-played log representing what actually aired to the traffic system which uses the data to expedite the billing process. In short, to be a good automation system, it is not necessary to possess scheduling features, but a good system must be able to use the schedules generated by other applications and then report back to them the on-air as-played results.

A good automation system will include several options for controlling when and how the system automates from event to event. Features to preset and/or execute transitions at a specific time of day are important for cueing up *stopsets* and starting programs. Transitions that tie the end of one event to the beginning of the next are useful for connecting multiple spots in a single stopset. Advanced systems often provide the ability to overlap events to provide segues between audio elements and to add voiceover announcements over background audio. The ability to perform these overlaps with dynamic level control (crossfades, ducking) all within a single stereo playback channel is highly desirable.

Another desirable automation system feature is the ability to record into the device while simultaneously playing audio. This allows production personnel to create new inventory without having to wait for specific breaks in the normal on-air programming. Most systems provide this function by offering multiple record and/or play audio channels. Some systems are scalable, allowing the channels to be added as needed by inserting additional DSP boards.

An important extension of this function is the ability to play a file while it is still being recorded. One excellent use of this feature is to perform short interval time shifts.

Once you begin recording a program, you can start playback of the same cut at any time, even before the program (and recording) is completed. This allows hour long programs to be delayed by as little as a few minutes. Another powerful use is to be able to play excerpts of a news feed while it is still being recorded. Sound bites can be used on air before the feed is over without having to make multiple progressive length recordings. Most automation systems require that a recording be in progress for a specified amount of time before the cut may be accessed for playback, but this interval is typically only a few seconds making this feature very handy.

An automation system rarely contains everything necessary for complete walk-away operation. For it to be truly effective, a system needs to be able to interface with various other broadcast devices, like—audio switchers, audio consoles, satellite receivers, CD players, DAT players, station clocks and more. These interfaces are created in two parts: the hardware interface, which provides proper electronic interconnection, and the software interface which communicates specific commands to/from the peripheral.

Systems based on personal computers can take advantage of a multitude of *off the shelf* hardware (plug-in I/O boards) that can provide any mix of RS-232, RS-422, RS-485 and large numbers of opto-isolated contact closure inputs and dry contact outputs. Systems that are able to address these *generic* I/O products are generally more flexible than those that offer their own custom I/O hardware that may offer a mix of these features.

But the greatest strength of an automation system is in its software interface—how it uses the hardware to communicate with the various devices external to the PC. A wide variation of command syntax and order exists among the various peripheral devices an automation system might connect to. Command strings, string lengths and the use of carriage return and/or line feed varies from device to device. Even simple contact closure controlled devices can often require specific closure sequences to perform a given function. It is therefore very important that the automation system software support the creation and transmisthat a more sophisticated "CD quality" version of the cart machine was needed to play spots, IDs, jingles etc.

As *digital cart* machines began appearing as direct hardware replacements for the existing analog units, some clever computer people were realizing that everything needed for a digital cart machine was already in a PC except the analog to digital and digital to analog conversion electronics. And this could easily be added by building a custom signal processing circuit board designed to plug into one of the existing PC expansion slots. Since PCs were already driving the development of the hard disk storage technology, they had access to the capacity required to store many hours of digital audio. This could be divided into any number of any size pieces, allowing an extensive inventory; and unlike tape or floppy disk based systems, PC based systems could provide *random* access to any and all of the material they contained.

Adding these powerful audio capabilities to the automation strengths of the PC has created the perfect platform for the continued evolution of station automation.

Key Features of an Automation System

The personal computer is capable of providing all of the key functions required for a powerful station automation system in a single box: random access CD quality audio, database operations, communications capabilities, programmability, and user-friendly operator interface. But to be an effective automation system, the hardware and software must be designed to provide certain key features.

Let's start with *CD quality* audio. Once the audio is digitized, it is manipulated as data, which the PC handles nicely. Therefore, the primary factor that determines the audio quality of a PC based digital audio system is the quality of the electronics used to convert between the analog and digital worlds. In a PC based system, this usually means the quality of the plug in digital signal processing (DSP) board. These boards include the analog input stage, the input low-pass filter, the analog to digital (A/D) and digital to analog (D/A) converter devices, the output reconstruction filter, and the analog output stage for each audio channel. It is very important that the analog stages have very low noise, wide dynamic range and be very linear. The input low-pass filter must have a steep enough response to preserve the maximum high frequency information for a given sample rate, should have minimal passband ripple, minimal ringing, and have a linear phase response. The analog to digital converter must have very accurate quantization levels (high linearity) and must complete each conversion very quickly (short conversion time). The digital to analog converter and reconstruction filter should use higher resolution (more bits) than the A/D to keep from raising the noise floor and limiting the headroom of the output signal.

It is often difficult to obtain many of these parameters directly without extensive technical documentation, which makes comparisons challenging. To complicate things further, many manufacturers will measure these parameters and report the resulting specifications using differing terminology and references. Without going too far into the specification jungle, we can state that most *professional* PC based digital audio systems use premium quality electronics with at least 16 bit quantization and therefore have similar excellent audio quality specifications:

Frequency response @ 48 kHz sampling 20 Hz – 20 kHz = ±0.5 dB

Dynamic Range (measured from digital clip) >90 dB

Signal to Noise (measured from digital clip) >90 dB

THD + N (measured at max level) <0.05%

Phase error between L–R channels <1 degree @ kHz

There are also systems available that utilize consumer grade multimedia audio cards. Consumer units typically have more high frequency roll-off for a given sample rate, more passband ripple, more interchannel phase error and higher distortion due to less expensive A/D and D/A converters. Generally speaking, products intended for the professional market (radio, TV, recording studio) use high quality components and provide noticeably better audio quality than consumer grade multimedia units. Pro units also usually provide balanced audio connections and direct digital audio connections where consumer units do not. And most pro units provide onboard DSP *horsepower* to perform real-time data compression and audio effects where consumer units do not. The bottom line is when it comes to audio quality, you get what you pay for.

With the ability to record and play high quality audio established, the next most important feature an automation system must possess is a way to organize its inventory. This is where the PC's database capabilities are utilized. The system should be able to hold numerous fields of information for each piece of audio recorded into the system. Things like the title, length, out cue, record date, start date and kill date should be stored along with a unique cut identifier (numeric or alpha-numeric) for every recording. The system should be able to sort the database by any of these fields, and should offer search functions able to find a string from any of the fields. It should be easy to change, update or remove data as well as to generate electronic or hard copy reports of the contents. More advanced systems might also include features like storing segue points, auxiliary cue points, back timer points, number of plays counters and the date and time of each cut's last use. Most importantly, the information and tools to manipulate it must be presented clearly and logically to the user.

From this overall database of the complete system inventory, an automation system must be able to create (or import from an external source) sub-databases that include the elements that will be used (played) in secession. These *playlists* might contain elements for a single stop set or might contain an entire day's pro-

3.6 STATION AUTOMATION AND NETWORKING

DAVID T. TURNER
ENCO SYSTEMS, INC., FARMINGTON HILLS, MI

INTRODUCTION

The term *station automation* is a very broad concept. It generally refers to the use of devices, processes and system interconnections designed to make a broadcast station (radio or TV) run a series of scheduled events automatically—without operator intervention. This includes every scheme ever devised from massive reel-to-reel tapes and interconnected cart machines to all encompassing control systems that steer the directional array, turn on the transmitter, play the audio, control the console crossfades and generate transmitter logs and billing information.

While the individual pieces of hardware involved in broadcasting have evolved to include powerful automatic functions and interconnect capabilities, nothing has advanced the station automation state of the art more than the personal computer.

SYSTEM OVERVIEW

It wasn't too long ago that station automation referred to a 15 in. reel of tape playing at 1 and 7/8 in. per second on an auto-reversing tape deck. This automation arrangement could provide most of a day's programming with minimal operator intervention, but the audio quality was mediocre due to the slow tape speed, and it required production personnel to invest a great deal of time assembling the contents of the tapes. Once the tapes were assembled, it was very difficult to make any changes, especially if the tape was playing on-air.

A more flexible arrangement used a stack of cart machines with the secondary cue (SEC) output of one wired to the start input of the next, and so on with the last machine connected back to the first. This allowed the operator to load several songs and commercials to play automatically in sequence, freeing up short amounts of time for other important operations. This crude form of automation was followed by numerous variations of mechanical beasts designed to hold and cycle dozens and sometimes hundreds of carts through multiple playback decks. These electro-mechanical marvels were capable of automatically sequencing days worth of programming.

These machines were incredible achievements in station automation and have faithfully served the broadcast industry for many years. However, their dependency on extensive mechanical transports made them high maintenance devices and limited their flexibility. Even the audio quality was hard to maintain due to the number of capstans and tape heads involved. Broadcasters needed newer, higher quality, easier to maintain, more versatile automation systems.

As CD players and DAT tape machines became common place in the home and in the studio, the audience demand for *CD Quality* broadcast audio also grew. Along with the improved audio this new generation of equipment brought improved control features that allowed broadcasters to elevate automation to a new level. In addition to the standard START and STOP functions, a much more complete set of instructions including shuttling and indexing were now available, usually through a serial data protocol. This enabled a single electronic controller to *talk* to multiple devices, directing them to queue up then play individual tracks from CDs and tapes containing multiple tracks. With *CD jukebox* devices available that could hold and play tracks from over 300 CDs and DAT tape machines that could queue and play hours worth of programming, a station could now automate for days at a time, if the controller was programmed properly.

While several manufacturers developed automation systems for their own equipment, these were mostly proprietary hardware and not able to communicate with devices from other manufacturers. These units often employed special keypads for data entry and ran ROM based programs that were difficult to upgrade or modify. This is where the Personal Computer made its biggest mark on station automation. PCs were already being used in business and at home to run database applications, perfect for storing a *log* of scheduled events. PCs had excellent serial (and parallel) communications capability and could be configured/programmed to *speak* any protocol required. PCs used a standard user interface (keyboard and monitor). PC based programs were stored on disk and ran from RAM and were therefore relatively simple to modify and reload. The personal computer was an excellent platform for station automation controllers.

A number of manufacturers created custom software packages using standard PC hardware to automate CD jukeboxes, DAT players, and fire relays to play standard cart machines. These were quite effective and could provide a good level of *walk away* automation, but some of the stations most important material, (commercials, IDs, and promos) were left to the weakest link—the cart machine. It became increasingly obvious names emerge and some industry leaders fade from vogue as they fall behind the pace of changing technology. Pressure is on broadcast engineers also to educate themselves in order to provide a competitive product to their listeners and viewers.

Special thanks to Elaine Jones of Comrex Corporation, Dave Burns of Harris Corporation's Broadcast Division and Daniel Queen of the Audio Engineering Society for their contributions, clarifications and assistance in preparation of this chapter.

BIBLIOGRAPHY

Chinn, R., Symetrix 601 Owner's Manual, Symetrix, Inc., Lynnwood, WA, 1992.

Eger, D. (Chairman), AES3-1992 (ANSI S4.40-1992), Audio Engineering Society, Inc., New York, 1992.

Ennes, H., *AM-FM Broadcasting,* Howard W. Sams & Co., New York, 1974.

Finger, R. (Chairman), AES-3id-1995, Audio Engineering Society, Inc., New York, 1995.

Mendenhall, G., Maintaining a 100% Digital Path From The Studio To The "On-Air" Signal, Harris Corporation, Broadcast Division, Quincy, IL, 1995.

Meyer, C., and Watkinson, J., *The Book: An Engineer's Guide To Digital Transmission,* NVISION, Nevada City, CA, 1997.

Noah, J., "Performing QC on Digital Audio," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, April 1996.

Pizzi, S., "Understanding Audio Data Compression," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, May 1996.

Skelton, T., "Audio for the Wide Screen," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, February, 1996.

Technical staff, A Primer: Digital Aural Studio To Transmitter Links, TFT, Inc., Santa Clara, CA, 1994.

Anonymous, "About Digital Audio Codecs," In Product Summary, Comrex Corporation, Action, MA, 1997.

equalizer with slider controls creates a visual picture of the frequency response curve that is being produced. Non-technical people find it easy to use because of this design.

The *parametric equalizer* provides more versatility in the hands of a trained operator. A specific frequency can be dialed in with a parametric equalizer for elimination or boosting. A graphic equalizer does not allow this frequency specific accuracy. The parametric equalizer also allows the adjustment of filter bandwidth or selectivity. A narrow bandwidth can notch out a hum or buzz. A wide bandwidth produces broad curve to boost or cut an entire band of frequencies. The typical parametric equalizer offers three or four sections covering the entire audio spectrum. Parametrics are prized as valuable tools in the production studio.

Microphone Processors

Stations spend thousands of dollars for processing equipment to compress and equalize their music program material. The studio microphone often suffers from processing neglect although it deserves major attention. Compression, equalization and de-essing of the studio microphone give the on-air talent a fighting chance of competing with the professionally produced program material played on the air. A good microphone processor elevates the local disk jockey closer to the voice quality level of professional voice talent.

Compression reduces the dynamic range of the natural voice providing more power and punch by raising its average energy; it becomes louder. The added power prevents the voice from being buried by song introductions and music beds. The equalizer provides a means of boosting regions of the voice spectrum that lack natural presence. Male jocks always want a generous boost of low-frequency energy. A muddy voice benefits from a boost in brightness from the equalizer's upper mid-range (around 2.5 kHz).

Sibilance problems show up at frequencies between 6 and 7 kHz. The de-esser monitors the energy level in this range and kicks in additional compression to reduce sibilance problems. When adjusted properly, the de-esser will take the edge off a sibilance problem without punishment to the high-end frequency response. A de-esser circuit requires fast attack and release timing and a narrow (less than 0.5 octave) bandwidth.

Without voice processing on the studio microphone, the station may find itself over-processing its music while stretching to achieve a suitable amount of processing for the on-air voices. The addition of a microphone processor provides a more balanced processing mix from on-air voices and produced program material.

Noise Reduction Systems

Noise reduction systems minimize source noise from analog audio tape, vinyl records and remote pickup (RPU) radio links. Dolby and dual-ended systems record encoded audio on tape and decode it to reproduce the original audio with a lower level of system noise.

Audio *companding* (compression/expanding) systems provide another option. During the recording process the dynamic range of the program material is severely compressed keeping it further above the noise floor of the recording medium. The process reverses during playback when the recorded audio is expanded restoring the normal dynamic range. These dual-ended systems provide benefit in audio tape recording but require identical equipment on each end of the record/reproduce process.

Dolby and companding noise reduction systems paint over the noise. Single-ended systems provide freedom from the encode/decode process by sweeping noise under the rug. These frequency sensitive gating devices eliminate all audio in certain bands when the signal falls below a fixed threshold. The theory of operation assumes that anything below the threshold must be noise and should be eliminated.

Single-ended systems offer a better solution than cutting the frequencies associated with tape hiss with an equalizer. The equalizer also eliminates desirable program material in that spectrum. The noise reduction unit attenuates a part of the audio spectrum only when no significant program material appears in that band.

Effects Generators

Digital effects generators enable production directors to produce the special effects of *phasing, flanging* and *echo* for creative production. Several effects devices allow easy pitch changes of voice and music for special effects and to compensate for speed adjustments made to time tapes perfectly to 30 or 60 seconds. When tape speed varies more than 2 or 3%, pitch correction restores normal tone to the voice.

When using reverberation in program material, the amount mixed in should be 20 dB down from normal program level. Otherwise the effect will be too distracting. When used with compression or limiting, it should be mixed back into the audio chain at the limiter output. This prevents the percentage of the mix from varying with the operation of the compressor/limiter.

SUMMARY

The standards that have served us well since the infancy of broadcasting are now falling to new technology. The digital AES/EBU (AES-3) interface is superseding the standard 600 Ω, +4 dB connection between audio equipment. Our long familiar hardware friends, punch blocks, screw barrier strips and XLR connectors make way for DB25s, fiber optic cable and ISDN.

The broadcast engineer can now build a studio using source material from digital audio storage devices mixed on a totally digital console. The console can even be a touch-screen controlled computer. The all-digital processing system has arrived. Some stations have already replaced their 950 MHz STLs with T1 carrier systems. Digital stereo generators arrived several years ago as did digital FM exciters.

Retooling to produce digital studio equipment is challenging the equipment industry. We have seen new is less apparent as the compressor begins to act on the rising audio signal. This describes a *soft-knee compressor.* The old tube-type compressors, the UREI LA-2 for example, remain popular because of the smooth transition through their soft-knee.

The combination of input level, threshold, and ratio set the amount of compression applied to the audio. The compressor's attack time and release time also factor into the proper setup of a compressor. The attack time specifies how quickly the compressor responds to a rising audio voltage as it crosses the threshold. A slow attack time, one measured in milliseconds, lets a few overshoots through. Aggressive compression utilizes a fast attack time, one measured in microseconds.

Release time determines how quickly the compressor releases the compression when a falling voltage drops below the threshold. A slow release time may measure 2.5 seconds or longer. A fast release time, measured in milliseconds, will release as each low frequency waveform decays. A high compression ratio coupled with fast release time may result in too much of a good thing.

When more aggressive compression hammers the dynamic range the resulting audio product becomes grunge music or punk rock. Gone are the normal soft passages in music, and normal voice migrates to a stream of shouted commands. Over-compressed audio becomes irritating over a period of time. Some musical notes can actually be lost. If guitar notes and the beats of a kick drum arrive at the input of a compressor at the same time, the compression required to tame the kick drum will drop the level of the guitar so much that some notes will not be heard.

Limiters

Take everything that you just learned about compressors and apply a 20:1 compression ratio; now you have a peak limiter. The limiter fits in the program chain just ahead of the transmitter (the stereo generator for FM). The limiter provides a brick wall for any overshoots which the compressor misses. Its sole purpose is to prevent overmodulation of the transmitter.

Limiters used for FM broadcast in the United States include the 75 msec preemphasis curve that boosts the high frequencies. Some countries outside the United States use other preemphasis curves; check the local standards. FM limiters operate symmetrically; positive and negative peaks receive equal limiting. AM limiters operate asymmetrically. They clamp their negative peaks at 100% modulation while allowing positive peaks to shoot upward to 125%. This maximizes the modulation and output power of the AM transmitter. AM limiters also include the NRSC preemphasis curve and 10 kHz low-pass filter. C-QUAM, AM stereo processing, usually requires an optional circuit board.

Most engineers locate the peak limiter at the transmitter site, feeding its input from the STL output. Leaving the compressor at the studio end maximizes the modulation capabilities of a 950 MHz STL.

AGC/Levelers

A novice may confuse an automatic gain control (AGC) amplifier with a compressor. The AGC/leveler is not an audio processing device. It will not reduce the dynamic range or limit loud audio peaks. The AGC amplifier insures that the compressor and limiter receive constant input levels.

The AGC operates like a compressor when a signal above its target output range appears at the input. In this case the AGC amplifier gently pulls the signal level down to the target output level. It is operating like a compressor set with a 2:1 ratio and a slow, two minute release time.

The AGC amplifier reacts differently when the input signal falls below the target output level. Now it turns up the gain of the low-level signal bringing it slowly up to the target output range.

AGC/levelers provide a safety net for sloppy board operation, the times when the meters remain buried in the red for minutes at a time. Stations employing walk-away operations, where no one sits at the console correcting levels as the automation system switches between sources, benefit greatly from the processing consistency created by an AGC amplifier.

Equalization

Engineers work nights in anguish maintaining audio frequency response as flat or linear as possible. Program directors toil overtime in anguish producing a signature sound for the station. The two professionals collide at the equalizer. In the audio processing chain, creative use of an equalizer gives character to the station's audio. In production, an equalizer routinely cleans hum and hiss from noisy tapes. The production director also creates special effects and unique voice tracks with the help of an equalizer.

Low frequencies can be boosted to produce a heavy thumping bass and the upper mid-range can be boosted to add brightness. Enhancement of the highs and lows with an equalizer satisfies the desire to make your station sound better to the average non-technical listener than the station across town. Table 3.5-2 matches the audio ranges and qualities with the knobs on your equalizer.

Equalizers come in two types. The best known, the *graphic equalizer,* divides the audio spectrum into a series of bands represented by rotary or slider controls on the front panel. The operator adjusts the controls to affect gain in each particular band. The graphic

Table 3.5-2
Common Audio Frequency Classifications

Range	Audio Frequencies
Sub-Bass	15 Hz–65 Hz
Bass	65 Hz–256 Hz
Voice	256 Hz–2048 Hz
Upper-Vocal	2048 Hz–3750 Hz
Presence	3750 Hz–5000 Hz
Sibilance	6000 Hz–7000 Hz
Brilliance	6500 Hz–15 kHz

to-reel recorders. One large system offers simultaneous access to program audio by seven studios. Digital audio storage systems remove the classification cart librarian from the on-air talent's job description.

Digital Editing and Work Stations

The digital audio workstation benefits the broadcast station with faster and more creative production, using a process quite similar to the way word processors edit written text. Digital workstation systems store multiple audio tracks on hard disk and allow editing in random access memory (RAM). Two or more analog input channels (through A/D converters) and direct digital inputs receive the incoming audio. Options allow analog or digital outputs. They allow editing tracks individually and produce a finished product by using keyboard and scrub wheel rather than grease pencil and razor blade. The audio waveform of each track crawls across a screen allowing visual as well as audible cueing and editing.

To produce a spot with a digital audio workstation, the operator records the audio tracks and music beds in the system memory, commands the workstation to move component sounds, adjusts timing, edits tracks and finally completes a stereo mix. Unlike razor blade editing with analog tape, the software process preserves the original material. Correcting mistakes and editing experiments become child's play with the undo key found on most editors.

Workstation editors most closely emulating the operation of a reel-to-reel recorder get production work flowing quickly with minimal training time. An accurate scrub wheel operation speeds the work of a producer tightening loose voice tracks and other tricky edits. Look for scrub audio that sounds exactly like that heard when manually rocking tape back and forth across the playback head of an analog tape deck.

A good display shows the audio waveform of all tracks on the screen simultaneously. Most editors show a vertical cursor line that moves across a stationary audio waveform as the editor reproduces the audio. This display mode consumes less computer resources than a system that shows the audio waveform crawling across a stationary cursor because the screen is not continually refreshed. Some audio producers favor a display that depicts the waveform moving across a stationary cursor as the screen more accurately represents the operation of a reel-to-reel machine where the tape travels left to right across the stationary playback head.

Creative engineers can assemble a two-track audio workstation with a Pentium PC, a good sound card and editing software. The system should include as large a hard drive as possible; the source material plus overhead consumes about 12 mB per minute for stereo audio. Some audio clips will certainly be archived on the hard disk so it is very important to get the largest one possible. More elaborate systems include mixing capabilities providing an all-in-one approach for production work.

AUDIO PROCESSING EQUIPMENT

The difference between the loudest sound and the quietest equals the dynamic range. Audio compressors take large, rapidly rising audio voltages and make medium-sized, slowly rising audio voltages. The goal remains simple; reduce the dynamic range and increase the average modulation level of the transmitter.

If the output of a CD player was directly connected to the audio inputs of a transmitter the wide dynamic range would prevent the average modulation from reaching more than 50 to 60%. The station would disappear from the band among those processing their audio by today's aggressive standards. An AM station operating in this manner would sacrifice a part of its coverage area. An audio compressor reduces the dynamic range to a more practical spread and holds the modulation level at a much higher value.

Here's how it works. A threshold control defines the point where the compressor starts to attack a rising audio voltage. If the compressor's threshold control instructs it to attack a voltage at −20 dB, the compressor will attempt to hold down a rising audio signal after it rises above the −20 dB level. A signal below −20 dB will pass through the compressor unscathed.

Once the audio passes the threshold (−20 dB), the ratio control instructs the compressor how serious it is about restricting the level of the audio voltage. A ratio of five to one (5:1), limits an audio signal which has risen five decibels to an increase of only one decibel at the compressor output. Using this example an input signal of −20 dB would appear at the compressor output at −20 dB; no compression would take place. An input signal of −15 dB (5 dB higher), would show up at the output at a level of −19 dB. (1 dB higher).

Low ratios induce less compression for any given input above the threshold. When the user increases the ratio, the compressor aggressively attacks the dynamic range of the source material. The threshold presents more of a "brick wall" to the incoming audio as the operator dials in ratios of 10:1 and higher. Premium quality compressors offer two thresholds with an independent ratio control attached to each threshold.

This type of compressor moderately squeezes a rising signal of low level as it crosses the first threshold level. Should the signal continue to rise above the second threshold, more aggressive compression attacks the signal. A compression ratio of 20:1 used above the second threshold provides the peak-limiting required to properly modulate a transmitter.

The compressor's *knee* is the point where the incoming audio signal rises above the threshold and the compression action begins. A *hard-knee compressor* engages the compression precisely at the exact threshold value with the exact ratio. A good ear hears the full amount of processing kick in as the signal rises above the threshold. If the compressor smoothes the transition with a gradual, rounded transition point at the threshold, the processing action engages slowly over about a 6 dB range. The transition into processing

System designers greatly increase the storage capacity of hard disks by using the bit rate reduction techniques of ISO/MPEG, Audio Processing Technology's apt-X, Musicam, the international telephone ITU G.722 Standard, Dolby's AC-2 or AC-3. Bit rate reduction, or audio data compression, allows six gallons of audio to be stored in a one gallon audio bucket. Without bit rate reduction one minute of stereo audio consumes about 10 mB of hard drive space. With a 4:1 bit rate reduction the same 10 mB could store four minutes of audio.

Early algorithms accomplished bit rate reduction by transmitting only the difference between samples. These systems gained a 2:1 reduction in digital data, but were non-destructive. The algorithm restored the original audio signal when decoded.

Lossy bit rate reduction discards bits not needed because the human ear would not hear the sound reproduced by the bit when played back. The sound would either be too low in volume to be heard, or covered up by a louder sound. Digital audio compressed in this manner, when converted back to analog, produces an audio signal almost indistinguishable from the original to the critical ear. This destructive process changes the source audio forever. Once compressed, the exact original audio can never be recovered. Huge gains in storage space and transmission bandwidth requirements justify the use of these algorithms.

The Moving Pictures Experts Group (MPEG) compression algorithm is the audio portion of a video compression system standardized by the International Standards Organization (ISO). The fact that it is accepted worldwide and is attached to a video compression scheme ensure that it will be with us for a long time. MPEG offers three levels of signal quality identified as Audio Layers I, II, and III. The complexity of the system, including hardware, and quality of the audio improves with each advancing layer. Audio Layer I provides transparent 20 kHz audio quality at a compression ratio of 4:1. Audio Layer II achieves the same results but at a ratio of 6:1. Audio Layer III operates as high as 24:1.

Musicam USA's compression system is compatible with MPEG Audio Layer II. Musicam chops the 20 kHz audio spectrum into bands of 750 kHz. It then discards any unnecessary bits in each band. The Musicam system represents a compromise between transparent recreation of compressed audio and complexity of the processing algorithm. Processing delays increase with complexity of any system's algorithm.

Bit rate reduction serves to allow digitized audio transmission over limited bandwidth mediums. Early digital STL systems used bit rate reduction to squeeze two channels of 15 kHz audio and remote control data into the 500 kHz bandwidths of the 950 MHz spectrum. Remember that bit rate reduction algorithms throw away part of the digitized audio, preventing further audio processing. All audio compression and limiting must happen before the program audio passes through a digital STL that uses bit rate reduction. There are, however, some digital STL systems that do not use audio compression algorithms. Such systems do permit audio processing to occur at the transmitter site, if desired.

T1 digital telephone systems offer enough bandwidth to allow uncompressed audio to reach the transmitter. A T1 STL system offers flexibility in the location of audio processors. Multi-band compression may be applied at the studio and final limiting accomplished at the transmitter site. T1 systems also solve the problem of the congested 950 MHz spectrum.

Digital Audio Storage Systems

Digital audio storage is the most rapidly developing technology in the broadcast audio field. Its acceptance by the industry became apparent in December 1995 when Ampex and 3M announced their exit from the magnetic tape business. The magnitude of this development becomes evident after considering that Ampex pioneered magnetic tape recording in the United States. Digital recording mediums now include: magnetic tape, hard-drives, high density floppy disks and magnetooptical disks.

Designed for the consumer market, the rotating head digital audio tape (DAT) machine became widely accepted by broadcasters. This format uses a rotating head much like a video cassette recorder (VCR), and records on a cartridge tape. DAT cartridges offer recording times of 60, 90, and 120 minutes. Professional models feature fast cueing, instant starts, remote control and time-code compatibility. DAT machines offer broadcasters the ability to record live events, concerts, and network feeds in the digital domain without increasing noise and sacrificing dynamic range.

Stations using multi-track consoles for production work may take advantage of eight-track DAT recorders. Like their smaller cousins, the eight-track models use a rotating head but they record on VHS video tape cartridges. These DAT recorders excel in recording live concerts and storing multi-track production work.

Smaller facilities may opt for the digital cart machines coming on the market. These record to super high-density floppy disks, mini-disks or magnetooptical disks. They offer the advantages of low-noise digital audio reproduction, elimination of head alignment woes and maintenance associated with analog cart machines. However, the disk jockey still has to sort and slam plastic devices into slots and pushing start buttons.

If you plan a transition to digital, why require talent to handle the storage medium at all? Digital audio storage systems using hard drives as the storage medium threaten to oust cart machines and carts, of all types from the control room. Hard drive systems now offer the capability to store and play all of the station's commercial library plus its program material. Mirrored hard drives, (multiple hard drives containing the same information) provide the necessary crash protection. The systems can automatically record network feeds for delayed broadcast, eliminating the need for reel-

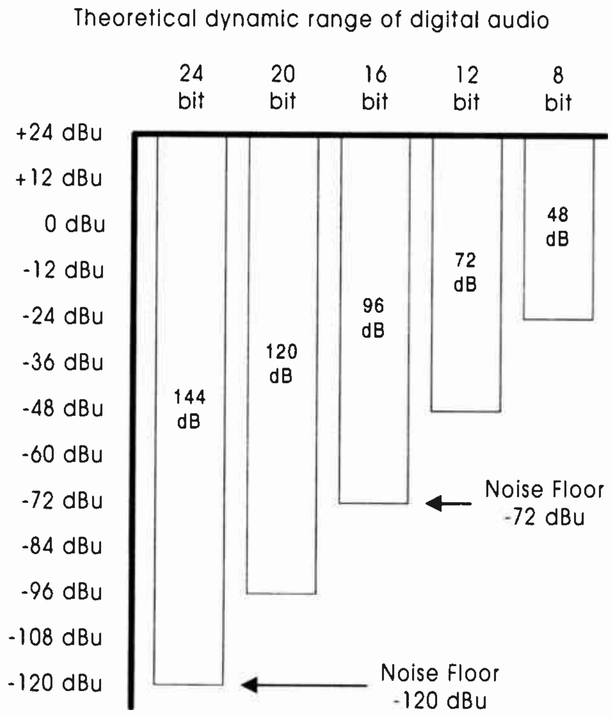


Figure 3.5-7. Higher quantization rates (bit rates) result in a wider dynamic range for digital audio. 24 bit audio provides a theoretical dynamic range of 144 dB for a 48 dB improvement over 16 bit audio. This also means a 48 dB lower noise floor.

An increase in dynamic range means a lower noise floor. Note the 48 dB reduction in the noise floor when comparing 16 bit digital audio with a 24 bit signal. Assuming a clip level of +24 dBu, 16 bit digital audio equipment can theoretically reproduce a minimum signal level of −72 dB. 20 bit digital audio has the potential of reproducing a minimum signal of −120 dBu. Figure 3.5-7 illustrates the difference in possible dynamic range between 24 bit, 20 bit and 16 bit digital audio.

AES3 standardizes sample rates of 32.0, 44.1 and 48.0 kHz. The digital recording process limits the maximum recordable audio frequency to one-half of the sample rate. Why? Each cycle of a sine wave must be sampled at least twice during a cycle. If not sampled once during its positive peak and once on its negative peak, the sound cannot be accurately sampled and converted to digital audio.

A sample rate of 44.1 kHz permits audio bandwidth up to 22 kHz. The sample rate of 48.0 kHz pushes the upper limit to 24 kHz. Equipment using a 32.0 kHz sample rate chokes off the audio at 16 kHz, but consumes less bandwidth and/or hard disk space.

Today, the existence of three sample rates hinders digital broadcasting. Compact discs came out of the gate with the sample rate of 44.1 kHz. This single standard sample rate crossed all brand names and entrenched itself as *the* sample rate for CD work. Recording studios adopted the sample rate of 48.0 kHz in an effort to achieve better fidelity. Digital audio workstation and digital audio storage equipment manufacturers shifted downward to 32.0 kHz to conserve disk space since FM radio only passes 15 kHz of audio bandwidth.

As digital audio consoles become more common in the control room, the engineer wishing to go completely digital in the studio has a problem. It is not possible to combine digital audio of different sample rates even though all of the sources adhere to the AES/EBU Standard. Somewhere in the studio system sample rate conversion must take place. The 44.1 kHz sample rate of the CD must switch to the 48.0 kHz sample rate of the console which must convert to the 32.0 kHz sample rate of the digital STL.

One console manufacturer offers input modules that match the sample rate of input audio to that of the console. Their solution to the problem is sample rate conversion on each input. A standard sample rate for all broadcast applications offers a better solution; but we are not there yet. Careful planning during equipment selection remains the best defense for minimizing potential problems.

Even after the sample rate issues are worked out, one more problem surfaces when a digital console arrives in the on-air studio. When mixing two digital audio sources of the same sample rate, their clocks must be synchronized. Suppose one digital audio source runs at 48,000.00 Hz and you mix it with a second signal running at 48,000.02 Hz. If the internal clocks run separately, digital train wrecks occur.

The results pollute the combined audio with clicks and pops. A digital master clock provides a workable solution. A single timing signal connected to all digital audio equipment in the station synchronizes everything in frequency as well as phase.

This plan works when all digital equipment in the studio offers clock input connections. Be aware, however, that not all digital equipment is designed with this feature. Older digital equipment featured a *word clock* that ran at the sample rate. Newer equipment locks internal clocks together with the AES Standard *Digital Audio Reference Signal* (DARS) which is an AES3 digital signal without audio. Check your digital equipment manuals to verify which systems will operate in your facility.

Three more pieces of the AES/EBU puzzle remain. The first bit of the digital signal indicates whether the output signal originates from *consumer equipment* or *professional gear*. If the first bit of the first 8 bit word is a 0, the source audio is from consumer equipment. Professional equipment is identified with a 1. Many professional recorders will not accept a signal from semi-pro equipment.

Digital audio may also include a consumer copyright protection bit. The Serial Copy Management System (SCMS) prevents illegal digital copying. Digital audio also includes a parity check-sum for the channel status data. The cyclic redundancy check (CRC) error light on your recorder may indicate a problem with your recording configuration.

POTS Codecs

The compression algorithms discussed previously were designed for use on Switched 56 or ISDN lines, which have a guaranteed data rate of 56 or 64/128 kbps. Standard analog Plain Old Telephone Service (POTS) provide varying data rates depending on line quality. Telephone modems are designed to scale the transmitted and received data rate accordingly. This sliding data rate makes it very difficult to use compression algorithms that were designed for specific transmission speeds.

In the last couple of years new technologies have emerged that permit the transmission of compressed audio over POTS, thus delivering a bandwidth much higher than the 3 kHz signal normally available. POTS codecs use high speed modems (33.6 kbps as of this writing) and compression rates in the 12:1 range to deliver audio bandwidths varying from 5–10 kHz, depending on the connection speed of the modems.

There is an important difference between audio sent by POTS codecs and audio sent via computer or the Internet. Computers use asynchronous modems, which means they send a packet of information, wait for confirmation that the packet was received, then go on to the next packet. For one-way delivery of audio that has been stored at a site, this method of transmission is not a problem. However, asynchronous modems must use a buffer to reconstruct the audio. This creates a delay large enough that two-way communication with the studio is simply not possible over the single telephone line used for the remote. For real time bi-directional remotes, most POTS codecs use synchronous modems. With a synchronous modem, a bit is transmitted, and without waiting for verification of receipt, the modem goes on to the next bit. This constant data stream in both directions results in audio that is delayed only by the compression algorithm used. The modems go through a complex *handshaking* process to determine the best data rate for transmission in order to minimize errors caused by lost bits.

DIGITAL AUDIO SYSTEMS

The Audio Engineering Society and European Broadcasting Union (AES/EBU) standards provide a benchmark for digital inputs and outputs. Fortunately, these standards provide a common denominator found on all professional equipment. AES/EBU, also known as AES3, created a balanced system that can transmit digital stereo audio up to 100 meters over a single shielded, twisted-pair wire. AES/EBU calls for internal transformer coupling, with dc blocking capacitors, on both input and output circuits. The circuit is designed for shielded audio cable with an impedance of 110 Ω.

Any digital input protocol must match the digital output of the source. AES3 provides assurance that any two pieces of equipment using this standard will work together.

The AES3-1992 Standard specifies a nominal signal voltage between 2 V and 7 V measured across a 110 Ω terminating resistor. An earlier standard (1985) allowed a 2 V to 10 V range. No compatibility problems exist between the 1985 and 1992 standards.

AES3 specifies connections using the familiar XLR three-pin audio connectors. Pins #2 and #3 carry the digital signal. Pin #1 is ground. This convention is exactly the same as used in wiring XLR connectors for microphone or balanced line-level audio. The connectors remain the same, but the wire does not.

Miniature broadcast audio cable (Belden 8451 or West Penn 291) does not meet the 110 Ω impedance specification. Remember, digital audio is actually computer data that runs at a rate of 64 Hz times the sample rate. Digital audio, sampling at a rate of 48 kHz, becomes a data stream running at 3.072 Mbps (64 Hz × 48 kHz = 3.072 MHz). Cable capacitance can rapidly degrade the 3 MHz signal. Belden, Gepco, Clark Wire & Cable and other cable manufacturers, offer low-capacitance, 110 Ω, digital audio cable that accomplish a satisfactory job of transferring the digital audio signal for the AES3 specified distance of 100 m (328 ft).

The need to transport digital audio further that 100 m resulted in a new standard using unbalanced coaxial cable. AES-3id-1995 allows transmission of a 1 V digital audio signal up to 1000 m (3280 ft). Considering the 3 MHz frequency of digital audio, 75 Ω coaxial cable makes a lot of sense. No knowledgeable engineer would suggest conducting a 3 MHz RF signal over twisted-pair audio cable. AES-3id calls for RG6A/U, or RG59B/U, cable with BNC connectors.

Television and video production facilities rapidly embraced this new standard. AES-3id not only allows for longer cable lengths, but permits the use of 75 Ω cable, terminated with BNC connectors, for both video and digital audio signals. You can expect the radio and studio engineers to retain AES3 and their trusty XLR connectors.

Sony and Phillips developed the Sony/Phillips Digital Interface Format (S/P DIF) a standard for consumer grade equipment. S/P DIF defines an unbalanced digital connection. S/P DIF and AES3 signals do not mix in all cases. An AES/EBU input will accept a S/P DIF output, but a S/P DIF input will not accept an AES/EBU output without an interface to correct the differences in the data formats and wiring.

S/P DIF specifies the old RCA phono pin connector. A 75 Ω coax works well as the conductor. The problem of yet another connector and cable type in the studio proves reason enough to standardize on AES3 digital equipment in the station.

The AES/EBU (AES3) Standard supports 16, 20 or 24 bit quantization formats. In theory, the 16 bit audio of a CD can communicate up to 96 dB of amplitude change (6 dB × 16 bits = 96 dB). 24 bit digital audio can reproduce a theoretical dynamic range of 144 dB (6 dB × 24 bits = 144 dB). In the real world, digital audio equipment achieves performance less than theoretically possible. 16 bit typically renders 90 dB of dynamic range due to the limitations of the analog to digital converters and digital to analog converters.

one or two of the on-air staff with a van full of prizes, amps, speakers and a microphone to Big Al's Used Car Lot. The primary concern is to transmit the voices of the talent from Big Al's back to the radio station with reasonable quality. The dial-up telephone network provides a cheap and easy solution.

The output of a simple microphone mixer, or small audio console, connected to a telephone line coupler provides the basis for the most elementary remote system. The person doing the remote calls the station, is connected to the control room console via the station's telephone hybrid, and monitors the off-air signal for cues. This system is easy to set up and operate but the dial-up phone system limits the audio quality. The telephone system, designed only to transmit voice from one telephone to another, limits the bandwidth to a range of 300–3200 Hz. The expensive air talent at the remote from Big Al's *sounds* like they called in on a telephone.

Telephone Frequency Extenders

The problem with using the dial-up telephone system for delivering broadcast audio is the limited frequency response. The fact that the telephone system rolls off all audio below 300 Hz costs 2.5 octaves of audio on the low end (50 Hz to 300 Hz). This tends to produce the tinny characteristic which makes unprocessed, dial-up, telephone remotes sound bad when compared to the full spectrum audio of regular programming. However, there is a solution.

Analog telephone frequency extenders trick the telephone system into passing audio with a bandwidth of 50–2900 Hz. An encoder-decoder process shifts audio frequencies upward by 250 Hz, sends the up-shifted audio over the phone line, then returns the audio to its normal frequencies on the receive end. The encoder travels to the location of the remote broadcast and serves the double duty of shifting the audio up by 250 Hz plus connecting the remote mixer to the dial-up telephone network. Most include telephone touchpads for dialing.

The encoder converts 50 Hz audio upward by 250 Hz to a frequency of 300 Hz. Audio at its natural frequency of 2950 Hz exits the encoder at a frequency of 3200 Hz which barely squeezes through the limited bandwidth of the phone system. The output of the frequency extender encoder sounds quite strange. Even the lowest baritone voice sounds like Donald Duck. Obviously, reverse treatment is required on the receiving end.

The frequency extender system's decoder shifts the audio which it receives down by 250 Hz, restoring it to the original frequencies. The analog frequency extension process delivers an audio bandwidth of 50–2950 Hz. The result is very pleasing voice transmission over a dial-up telephone circuit. The process sacrifices 1/7 of an octave between 2950 and 3200 Hz, but it restores 2.5 octaves between 50 and 300 Hz.

The system suffers from two disadvantages, First, analog frequency extension is a one-way system. When the remote is out of the range of the broadcast station's signal, a second telephone line is required for talkback and cueing. Second, compatible equipment is required on each end. A station in New York cannot send frequency extended audio to a station in Los Angeles unless both have identical equipment.

Digital Audio Codecs

The computer age has brought another means of transferring high-fidelity audio from one point to another. Broadcasters can now digitize the source audio, apply data compression techniques and send the audio over telephone lines via modems. Because computer modems operate bi-directionally, two-way audio communication can be achieved over a single phone line.

In brief, a digital audio codec (coder/decoder) consists of an analog-to-digital converter, a digital-to-analog converter, a modem and data compression software bundled into a single package. The modem is designed for Switched 56, ISDN or dial-up lines. Switched 56 service provides data transfer at a rate of 56 kbps. ISDN doubles the speed to 128 kbps.

Dial-up lines limit data transfer to no more than about 53 kbps on average and data rates will vary from line to line, with weather, and with telephone traffic conditions. 56 kb computer modems do not always connect at a speed of 56 kbps. Switched 56 and ISDN service are consistent. The trade-off for less audio bandwidth brings the ease of connection to any existing telephone line without extra line charges and construction delays. Four-hour commercial remotes make the use of a dial-up line an easy decision. Coverage of a week-long special event may justify the expense of an ISDN line.

Systems designed for Switched 56 lines provide a 7.0 kHz, bi-directional, audio circuit. ISDN service doubles the bandwidth to 14 kHz, or allows stereo 7.0 kHz audio transmission. Improvements in modem speed and technology now challenge these premium services with bi-directional audio bandwidth of up to 10 kHz over dial-up telephone lines. The program audio bandwidth capability not only depends upon the bandwidth of the telephone circuit, but with the data compression algorithm used in the codec.

The audio compression (data reduction) algorithms most frequently employed with Switched 56 and ISDN lines include ISO MPEG Layer II, ISO MPEG Layer III, apt-X, Musicam United States or the international telephone standard ITU G.722 (formerly, CCITT G.722). The rules of compatibility dictate that the codec on each end of the telephone line use the same algorithm. MPEG Layer III will operate with a Layer II device, but performance is limited to Layer II levels. MPEG is not compatible with ITU G.722. Likewise, apt-X only talks with apt-X. Musicam USA will converse with MPEG Layer II.

ITU G.722 introduces a minimal delay making it the most popular algorithm for talk shows and live remotes. Stations often find themselves using their top-of-the-line codecs in the ITU G.722 mode.

audio is canceled out by the out-of-phase wiring of secondaries #2. This leaves, in theory, only the receive audio, the caller's voice, on the primary of T2.

If the phone line exhibited a perfect 900 Ω resistive load, a 900 Ω resistor at the location of the equalization network would produce a perfect match. Telephone lines, however, do not represent a perfect world. Capacitance creeps into the mix because the phone company uses miles of #24 gauge twisted pair to connect to their central office. That much twisted wire forms a nice capacitor. The resistance of the circuit varies with the length of all that #24 gauge wire running back to the phone company. Loading coils, used in the telephone circuits to flatten frequency response add the inductance.

We connect our hybrid to what amounts to a RCL network when we plug into the phone line. The null created by the out-of-phase circuit created by the #2 secondaries no longer matches the amount of transmit audio induced into T2 through its #1 secondary. A tunable equalization network wired between the #2 secondaries compensates for the electrical characteristics of the phone line circuit. When improperly tuned, audible amounts of out-of-phase host audio mixes with the caller's voice making the host sound hollow. The hybrid circuit may cascade into feedback in extreme cases of mismatch.

An analog phone hybrid must be tuned for best performance for each phone line to which it will connect. The process involves transmitting pink noise during a call to an outside phone number then tuning the equalization network for a minimum level of pink noise at the hybrid's caller output. Digital phone hybrids accomplish this task with a short pink noise burst at the beginning of each phone call. The result is perfect separation of host and caller audio with excellent sounding talk show audio.

Telephone hybrid systems package multiple hybrids with call switching functions. The combined systems eliminate the challenge of building a talk show phone system from scratch with equipment from several manufacturers. Taking a system approach also solves the problem of creating multiple mix-minus feeds when using multiple hybrids to conference callers on the air. Time saved during installation usually justifies any extra cost of a combined hybrid and call switching system.

Broadcast Delay Units

Radio talk shows can be hazardous to a station's liability insurance. The spontaneity of a good talk show assure, that callers can, and will, say anything. This prompts the need for a system that allows time to pull the plug before certain words, or accusations, pass through the transmitter.

Before digital technology accomplished this task with no moving parts, there was tape. Tape delay systems employed a special cart machine with an erase head. A ten-second cart was inserted in the recorder. Real-time audio was recorded on the tape. It took ten seconds while the tape looped through the cart before reaching the playback head. The output audio emerged from the recorder ten seconds after it was recorded allowing time for the talk show producer to interrupt the delayed audio containing a profanity.

Simple digital delays perform this task without the worries of moving parts and broken tape. The device simply converts audio to a digital signal, records it in memory, then plays it back ten seconds later. The talk show producer mixes the program in the console's audition channel which feeds the input of the digital delay. The output of the delay routes to the program channel of the console and to the transmitter.

If things go wrong, the producer turns off the mixer carrying the delayed audio to the program channel and inserts fill music. The option also exists of switching the show host from the audition channel to the program channel after dropping the caller. The host then resumes the program in real-time.

Both tape delay and fixed-time digital delays pose the problem of transition in and out of delay. If the show host simply starts talking at the beginning of the program, their words will not exit the delay system until ten seconds later. Meanwhile, the audience is treated to ten seconds of silence. Most stations overcome this by playing a ten-second recorded introduction to the show on the program channel as the show host begins talking on the audition channel. The host's voice exits the delay precisely as the recorded message ends. If timed properly the transition is seamless.

A better approach to the problem of keeping talk show audio respectable is the digital delay unit that gradually builds the delay at the start of the program. This delay digitally records the real-time audio into memory and while it is building the delay time, it plays back the program audio slightly slower than it is being recorded. This process gradually fills the memory until the delay time reaches the maximum. After the memory fills, this delay operates just like the fixed-time delay; the audio appears at the output ten seconds after it is received at the input.

If the caller says something offensive, the host pushes a *dump* button which erases all, or part, of the audio in memory. Since the profanity was stored in memory, it disappears when the dump button is pushed. The program now is on the air in real-time and the delay begins the process of rebuilding the delay time again. After about a minute the host has enough delay in memory to begin taking callers on the air again.

This type of delay allows easy return to real-time at the end of the talk program. A few minutes before the show ends, the producer puts the delay in the *exit* mode. Now the delay records the program, but it plays back slightly faster than it is recording. This eventually depletes the audio stored in memory and the program returns to real-time. The best of these delays feature a relay bypass which takes the delay off-line when it is not in delay, or if it fails.

AUDIO REMOTE SYSTEMS

One of the most profitable activities for a radio station is the commercial remote. This consists of packing off

Reel-to-reel recorders operate by mixing the incoming audio with a high-frequency ac bias signal of fixed level and frequency. This combined signal magnetizes the tiny ferric oxide particles attached to the plastic tape as it moves past the record head. During playback, the play head converts the magnetic fields stored on the recorded tape to an audio voltage sonically equal to the signal originally recorded.

The bias signal assures that the record head creates a magnetic field sufficient to fully penetrate the ferric oxide portion of the audio tape. The frequency of the bias signal must be supersonic and typically is at least five times the frequency of the highest audio frequency recorded on the tape. The bias signal may be optimized for a particular audio tape by adjusting the bias level to produce minimum harmonic distortion when recording and reproducing a sine wave at a frequency in the range of 2–3 kHz. The operator's manual for the reel-to-reel recorder will contain instruction for optimization of the bias signal.

During playback, an equalization curve applied in the playback preamplifier insures that the reproduced audio produces a mirror image of the audio previously recorded. The equalization curve corrects inaccuracies related to the electrical characteristics of the record head, playback head and speed of the tape as it moves past the heads. Both low frequency and high frequency compensation perfect the playback process.

Proper alignment of both the record and playback heads is very important. Both heads must remain exactly perpendicular to the tape as it moves past them. Incorrect azimuth (side to side) alignment causes poor high frequency reproduction. Stereo recorders exhibit a loss of stereo separation when allowed to drift out of perfect 90° azimuth alignment. Improper zenith (front to back tilt) also contributes to the high frequency reproduction problem.

Professional machines include a third head, an erase head, in the tape recording process. This head uses the bias signal to clear any previously recorded audio from the tape during the recording process. If damaged, or not properly aligned, the erase head will leave remnants of audio beneath the new recording. In extreme cases, the old recording can be heard during silent, or low-level portions of the new recording.

In the radio station control room, the reel-to-reel records news feeds from networks and reporters in the field. FM stations record music requests and contest winners on a reel-to-reel for delayed playback. Some program material still arrives in the station on reel-to-reel tape. A two-track stereo deck with speeds of 7 1/2 IPS (inches per second) and 15 IPS fills the requirements of most control rooms.

Tape reels used to constantly turn in the typical production studio. Tape edit points, marked with a grease pencil, were cut with a razor blade and edited in a splicing block. Tape containing unwanted audio was discarded and the two edit points spliced together. Words were cut out and loose cues tightened using this method. This worked well with mono or two-track stereo formats. Nowadays, however, most radio studios either have already converted, or are in the process of converting to digital audio workstations. In many facilities reel-to-reel equipment is maintained only for backup purposes—or to retrieve archived material.

Telephone Hybrids

In many respects, the network between the telephone company's central office and the home, or business, remains largely the same as it was at the turn of the century. Broadcasters still depend on two copper wires to transfer voices from one place to another.

Both the voice being transmitted and the voice being received mingle back and forth on the same pair of wires. A telephone hybrid converts the two-wire system into a four-wire system which separates the caller's voice from that of the talk show host. Figure 3.5-6 shows the theory behind a telephone hybrid.

In this example, the core of the hybrid consists of two transformers, each having a single primary winding and two secondary windings. The talk show host's voice (transmit audio) feeds to the phone line from the primary of T1 and through secondary #1 of T1. The caller's voice (receive audio) travels from the phone line through secondary #1 of T2, then to the primary winding of T2. Note that the transmit audio also passes, in series, through secondary #1 of T2. This means that the transmit audio appears in T2.

Notice what is happening with secondary #2 on both T1 and T2; they are wired out-of-phase in relation to the first secondary windings. Although the transmit audio is introduced into T2 because secondaries #1 of both transformers are wired in series, the transmit

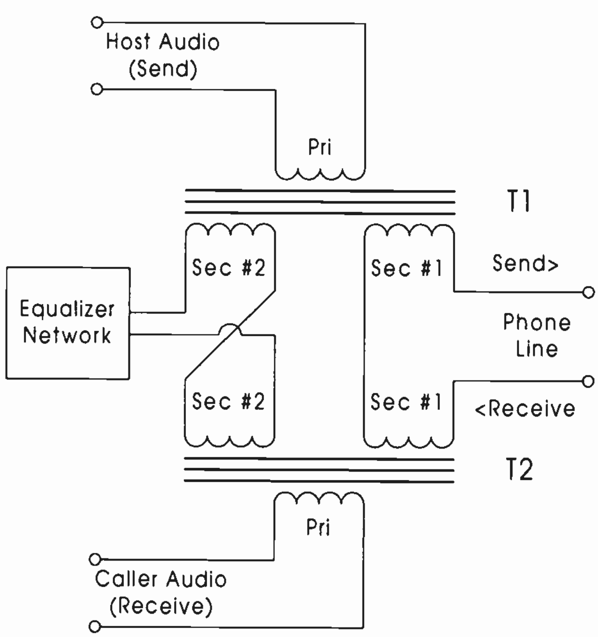


Figure 3.5-6. The telephone hybrid circuit converts the two-wire telephone line into a four-wire circuit. The hybrid creates individual send and receive audio connections separating the caller's voice from that of the host.

read by a laser beam focused on the spinning disc. Since nothing but the laser beam touches the disc, there is no wear.

Selecting the best equipment that the station can afford represents the best cost effective choice. Several manufacturers build CD players designed specifically for broadcast and professional use. If the station must use semi-pro players remember that their design considerations allowed for use in your living room a few hours a week. These consumer grade machines will not last indefinitely when run in a radio station 24 hours a day.

It is good practice to keep two spare (meaning new, unopened, in the box) semi-pro players in the station for quick replacement of a failed machine. Do not attempt to repair a failed consumer grade machine. They can be replaced with less hassle and expense than making repairs.

The output level and impedance of semi-pro CD players are not the same as their broadcast cousins. If your console inputs require +4 dB levels and present 600 Ω loads, use a matching interface. These matching boxes convert the −20 dB, high impedance, unbalanced output of the consumer grade CD player to a +4 dB, 600 Ω, balanced source. Consumer equipment can be difficult to cue and slow to start; evaluate units carefully before you commit to purchasing a quantity.

Turntables

Some unique source material remains available only on vinyl. This author disputed any remaining use of turntables in broadcast applications until the owner of one broadcast equipment manufacturing firm told me that he still sells 40–50 phono preamplifiers a month.

Turntables come in two varieties, the idler wheel design and direct drive. The once common broadcast turntable used a motor which turned at 1,800 RPM driving an idler wheel. The idler wheel in turn drove a large hub at the center of the platter. This design minimized wow and flutter caused by fluctuations in motor speed. The use of a heavy platter achieved further speed stability. Rapid starts necessary for tight cueing required a heavy, powerful motor.

The direct-drive turntable became more popular because of its reduced noise, wow and flutter. There is no idler wheel to replace or bearings to lubricate; the platter is the rotor of these slow turning electronic motors. The speed control circuits of direct-drive turntables constantly monitors and adjusts their speed, keeping it more accurate than if left to line voltage and frequency. Further contributing to the demise of the rim-drive tables, the speed control function made precise speed enhancement of music possible. Circuit repairs may be a problem because of the minimal documentation provided with most direct drive turntables.

No real broadcast tone arms remain; all current models are designed for consumer use. They track well and adjust easily but some prove difficult to cue and are not very rugged. A professional tone arm can be adjusted once and then left alone except for occasional testing. When installing a tone arm, use the template from the turntable manufacturer and follow the instructions supplied with the arm. Set the tracking weight as specified by the cartridge manufacturer.

The choice of the phono cartridge depends on the audio quality required. Rugged, less expensive models give the longest life in on-air use. Moving up to more expensive but less rugged models gains better separation and high frequency response. Avoid consumer grade phono cartridges.

Once mounted out of sight and never seen after installation, the turntable's preamplifier is easily neglected. The important specifications of noise, frequency response and separation need consideration. More expensive models offer filtering, high frequency cut or boost and adjustable cartridge loading.

Another pressing concern for the engineer is the preamp's resistance to radio frequency interference if the studio is co-located with the transmitter. RF easily makes itself known in these high gain amplifiers.

Audio Cart Machines

During the past forty years, cart machines proved invaluable for playing commercials, jingles and music. More importantly, cart machines removed the menace of 3 in. tape reels from the control room. Even stations that rely on digital audio storage systems will keep a few cart machines around as a backup system.

A mono machine utilizes two tracks, the upper track for program material and the lower for cueing. Stereo versions use three audio tracks on the endless loop of tape. Two of the tracks record stereo audio, the third carries cue tones. Trading recorded carts between mono and stereo players will not work because the tracks do not line up.

A brief 1000 Hz tone is recorded on the cue track at the beginning of the cart recording process. When the cart recorder is in the record mode, pushing the start button generates the 1000 Hz *stop tone* and begins the recording process. After the tape loop cycles through the cart and returns to the starting point, the playback head detects the 1000 Hz stop tone and stops the tape at the beginning of the recorded program material.

Most machines offer secondary and tertiary tones for cueing and starting the next tape. The secondary, or *aux tone*, is at 150 Hz tone and customarily triggers the next event in the program sequence of automation systems. The *tertiary tone* is an 8 kHz tone and triggers a cue light to warn air talent as the program material nears its end. The operator manually inserts the secondary and tertiary tones as the cart is recorded.

Although still a dependable and reasonably good storage medium for commercials and music, digital audio storage systems have moved essentially the cart machine out of the radio business.

Reel-To-Reel Tape Recorders

The reel-to-reel still remains a workhorse in some stations. Tape provides an economical means for storing longer program material without filling the hard drive of the station's digital audio storage system.

bility of a routing switcher provides the only practical solution when many audio sources must be switched frequently such as in a busy TV control room.

Size determines the cost of a routing switcher. A stereo switcher with 12 inputs and 12 outputs contains 288 cross-points (12 inputs × 12 outputs × 2 audio channels = 288). Visualize the switcher as two side-by-side matrices of 12 horizontal lines (inputs) intersected by 12 vertical lines (outputs). Each intersection becomes a possible connection point. One matrix represents left channel audio; the other represents the right channel.

In a television station, engineers must decide whether individual switchers handle left, right, mono, SAP, or if one large system routes all signals and audio. With the larger system approach, mono sources connect directly into left and right channels. The switcher may correct channel reversals and create mono mixes. A switcher with sufficient cross points to handle all switching tasks requires a larger investment than several smaller ones assigned to individual channels. The station planning for DTV is looking at a major investment in routing for the system's six channels of audio.

Distribution Amplifiers

When distributing audio to a number of locations on a continuous basis (without switching), a distribution amplifier (DA) proves invaluable. Sending a console's output to several recorders and other studios, or routing a satellite receiver's feed to all studios assure the distribution amplifier an important role in audio routing. A DA eliminates the need for constant patching and switching of various pieces of equipment. Distribution amplifiers represent the only practical solution when audio must be fed in multiple directions on a constant basis.

The typical DA provides six to eight stereo outputs for each channel. While there may be no input level adjustment, economical units should provide individual output trim pots. Modular distribution systems offer more versatility and avoid wasted, unused outputs. One model offers four stereo inputs, which can be assignable to any of its 14 stereo outputs by the use of jumpers. Popular DA options include metering, input level adjustment, audio compression, loss-of-signal alarms and redundant power supplies.

Studio Monitors

The control room audio monitoring system provides the first line of defense in spotting equipment failures and problems. For that reason, professional monitor speakers should be selected.

In choosing monitors, room size dictates cabinet size. In a large studio, invest in monitors with 12 in. woofers, 5 in. mid-range cones and horn or dome tweeters. Buy the best your budget will allow. Size limits small studios to a model with 5 or 6 in. woofers. Current speaker technology offers amazingly good sound quality from small cabinets. Look for low distortion and flat response.

A meticulous studio designer will consider background noise sources, reverberation time of the room, interaction from walls and ceiling and room equalization. Doing this properly means testing the control room with a real-time analyzer and positioning the monitors for best results. This is seldom practical.

When mounting the monitors on walls, use suspension mounts, preferably with vibration isolating components. Position each monitor an equal distance from the operator's normal position. Use sound proofing material on as much of the flat wall surfaces in the room as possible.

In small studios, *near-field* monitoring provides the best solution. Position the monitors in a triangular arrangement with equal distances between the monitors and the ears of the operator. Near-field monitoring assures that the monitors will be close enough to the listener that the direct audio from the speakers will overpower reflections and any undesirable acoustics of the room. Mounting solutions include a shelf, or wall brackets, above the console, suspension from the ceiling, or floor stands behind the console but directly in front of the operator. Position near-field monitors at, or just above, ear level.

The power amplifier becomes another vital consideration. Space and power requirements limit audio console internal monitor amplifiers to 10 W or less. Noise and distortion specifications may not be as good as those of stand-alone amplifiers.

Matching the power amplifier with the requirements of the monitors means another task in studio design. Pushing a low power amplifier to provide adequate listening levels can cause audio waveform clipping with distortion on peaks. Operating in this manner could damage the speakers. A better choice would be to operate a more powerful amplifier in a conservative manner.

To prevent rock-n-roll DJs from blowing the speaker voice coils with too much power, install fast-blowing fuses in the lines. Experiment with fuse values and listening levels to find the proper combination. Remember to stash a few spare fuses in the control room.

Just as important as amplifier power is the wiring between amplifier and monitors. Use at least #16 AWG for low-power amplifier and speaker combinations. Use heavier wire, up to #12 AWG, for combinations above 100 W or long runs of speaker wire. Consider that the cables from the amplifier, out of the rack, across the ceiling then down to the monitors may use 50 ft of wire even in a small studio. Audio purists insist that the wire length for both speakers remain equal.

AUDIO SOURCES

The CD player has been the audio source of choice in radio for nearly two decades. CD technology encodes audio as digital bits recorded as etched holes on the surface of the disc. A transparent plastic coating protects the surface so that only an accumulation of dirt or scratches affect the playback quality. The bits are rules still apply to these new boards. The only exception is that they pass along a digital signal rather than analog.

Audio Distribution and Routing

There are three basic types of patch panels (jack fields). The tip/sleeve 1/4 in. jack size is the oldest type, consisting of one conductor and one shield. This obsolete design dates back to the early days of radio. The tip/sleeve patch panel requires four single plug cords to patch a balanced stereo connection.

The tip/ring/sleeve 1/4 in. panel remains the most popular patch panel for radio. It offers two shielded conductors per cable. A pair of single-plug cords will complete a stereo circuit. Dual-plug cable assemblies allow the convenience of patching a stereo source with a single cable.

The most useful 1/4 in. patch panel design includes dual rows of 24 jacks. This configuration allows stereo pair spacing. These panels usually group their jacks in pairs with wider spacing between stereo pairs. This spacing technique, when used with a dual plug patch cord, makes it impossible to cross-patch an audio source. Cross-patching occurs when the user inserts the first single patch cord in one audio feed and the second patch cord in the adjacent audio feed on the patch panel. The dual plug will align only in paired jacks; cross-patching becomes impossible with these types of patch panels and dual plug cords.

Patch panel jack numbering follows the following convention: the jack in the top, left corner is identified as jack #1. Counting across and to the right, the last jack on the top row becomes jack #24. Jack #25 falls below jack #1 and is the first jack on the bottom row, starting on the left. Jack #48 is located on the bottom row at the right end and below jack #24. Although patch panels may contain more, or less, than 48 jacks, this numbering system remains the standard for identifying individual jacks.

Quarter inch patch panels also come in single rows of 26 jacks or dual rows of 52. These panels have standard spacing between all jacks and allow an additional stereo circuit on the 26 jack version and two additional stereo circuits on the 52 jack, dual row model.

Special configurations of the 1/4 in. jack field can create very useful designs. One version offers three rows of 26 jacks for a total of 78 jacks on the panel. The wiring scheme of the two lower rows create a conventional dual-row patch panel. Wiring the top row directly to the circuits of the middle row of jacks allows monitoring these equipment output circuits by patching between the top row jacks and the monitor amplifier inputs. Inserting a plug into the top row jacks does not interrupt the normal audio path through the patch panel.

Another custom item is a patch panel with special jacks that not only switch the conductors, but the shield also switches when a patch cord is inserted. Patching microphone circuits requires this seldom-used configuration.

One arrangement features patch panels built into a 19 in. rack mount chassis. The entire assembly mounts into the equipment rack just like the equipment that it connects. The jacks appear on the front of the rack and the rear termination points offer easy access to equipment wiring from the back of the rack. There remains a misconception that this design provides protection from RF. Do not count on it; the phenolic bay fronts provide no shielding of the jacks and some manufacturers even wire these designs with unshielded wire.

Rapidly making its way out of the recording studio and into broadcasting is the bantam or tiny-telephone jack field. These 0.175 in. in diameter plugs and jacks feature the tip/ring/sleeve configuration. The bantam patch panel consumes about half the space that a similar 1/4 in. jack panel would require in an equipment rack. Ninety-six jacks fit in a 1 3/4 in. by 19 in. rack space.

Patch Panel Wiring and Termination

Experienced engineers never connect patch panel jacks directly to equipment inputs and outputs. Termination blocks speed patch panel installation and offer flexibility when making wiring changes. Accepted practice mounts patch panel termination blocks in individual equipment racks, in a group in each studio, or in a central point in the engineering area.

Short jumpers between connections on the termination blocks complete the links between inputs and outputs. This allows wiring changes at a convenient, easy to reach location without pulling new wire between equipment. Terminations used include solder type *Christmas trees*, wire wrap, Type 66 telephone punch blocks or the newer types of punch blocks designed for stranded copper wire.

Christmas trees remain popular with engineers who trust only solder connections. They still serve well in remote trucks because of vibration concerns.

Broadcast engineers have switched to punch-type terminations. A specifically designed tool punches insulated wire into a slotted connector. The wire insulation strips away as the wire pinches into the connection. The process eliminates the soldering task associated with Christmas tree blocks. The "66" block accommodates #22 gauge solid wire used in telephone service. Using only solid #22 wire assures dependable connections. Stranded wire generally works but the stranded conductor may flatten out preventing removal of all insulation. Some strands may also break during the punch-down process.

Audio Routing Switchers

The audio routing switcher offers an alternative to patch panels. This unit accomplishes the same function by switching the audio with relays or solid-state switches rather than by plugs and jacks.

The system eliminates patch cords, can often be operated by remote control and may often route audio to more than one feed at a time. Some of the more elaborate systems feature computer control. The increased flexi-

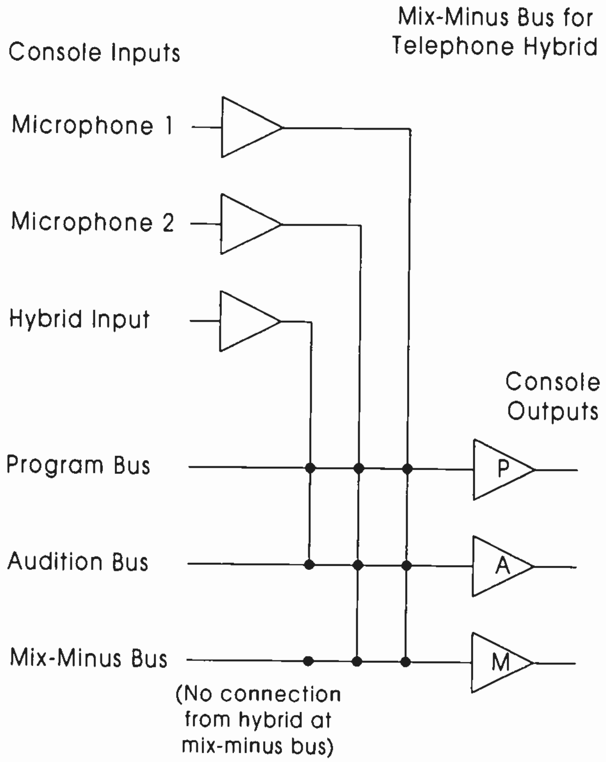


Figure 3.5-5. Broadcast consoles used with telephone hybrids require a mix-minus bus to feed the send input of the telephone hybrid. The mix-minus allows the caller to hear the talk show host's voice (Microphone #1 and Microphone #2) but prevents the caller's audio form being fed back into the hybrid.

1 and Input 2 (announcer microphones) connect to program, audition and mix-minus circuits.

If the studio contains multiple phone hybrids, a mix-minus feed for each hybrid is required so that callers will be able to hear each other's comments. In a television studio, a mix-minus feed provides on-air monitoring for talent on a live set. This feed includes program audio minus the talent's microphone preventing acoustic feedback while allowing the talent to hear program material and cues.

Consoles designed for radio production, and television, contain prefader processing patchpoints. These route the audio source compressors, equalizers or other signal processing devices before arriving at the mixing bus. These processing loops provide convenient connection points for microphone processing units.

Monitor amplifiers may be external or built into the console. Power and space limitations restrict built-in monitor amps to less than 10 W. Many engineers prefer to drive studio monitors with external, higher power amplifiers. The stereo monitor system should contain a stereo/mono switch. This allows the operator to check out-of-phase program material and misaligned tapes. A single-pole, single-throw switch wired between left and right channels at the monitor gain control accomplishes this function. At some stations, the control room monitors routinely operate in the mono mode. This immediately alerts station personnel to out of phase conditions.

A well designed console offers switchable input levels and impedance matching on each input module. This feature allows easy transfer of input sources to different mixers allowing for future changes in the studio. An input amplifier could then be used for either microphone or line level audio. A second best system would have interchangeable mixer input amplifiers which an engineer could shuffle between positions in the console mainframe.

Going one step further, some consoles offer programmable presets for input configuration. The engineer first stores pre-programmed console input settings in memory and then tells the console which program he plans to use. All of the input sources then automatically switch to the proper mixer resulting in fast and accurate setups.

Professional consoles feature balanced, bridging inputs. Bridging audio input circuits used in modern consoles handle a wider variety of input sources than possible when everything was designed with 600 Ω, +4 dBm terminations. A bridging input provides a 10,000 Ω, or greater, termination impedance which provides essentially no load to the source equipment. If an output requires a 600 Ω termination, a 620 Ω resistor tied across the input provides the necessary matching.

Remote start contacts for cart machines and other program sources became standard equipment in the 1980s. They allow the operator to start the equipment by simply turning on the appropriate channel. Some console manufacturers provide more flexibility by using logic circuits that allow the mixer to be turned on by pushing the start button on the cart player. When the cart machine recues the mixer automatically turns off. Look for automatic disabling of this logic when the input selector is switched to another input source. This eliminates the annoyance of having a cart machine start when the mixer is turned on for an auxiliary function.

Console manufacturers offer consoles with a choice of conventional analog VU meters or LED bar-graph metering. LED metering provides multiple color visual monitoring of root mean square (RMS) audio voltage plus peak values. One model shows left, right and peaks a single display. LED displays may make operators less likely to run a board with the meters buried in the red.

Multi-track recorders and digital workstations in the broadcast production environment require consoles with more than a single pair of left and right outputs. Four and eight channel consoles assist in producing award winning production. Channel assignment switches route the audio to the proper bus. Pan pots then shift it between left and right. Equalizers on each mixer allow adjustments to each audio source. Such production consoles resemble those once found only in recording studios.

Digital technology recently spread into the audio console industry in the mid-90s. The basic operating

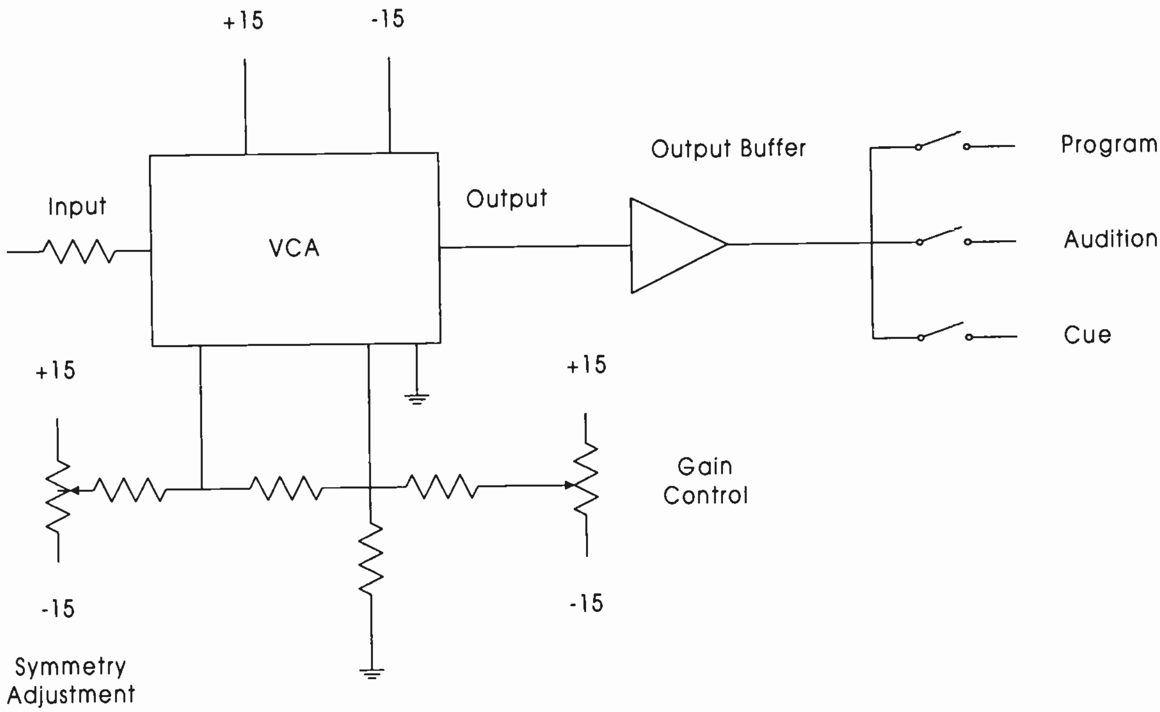


Figure 3.5-4. The voltage controlled amplifier (VCA) allows gain control via an adjustable dc voltage. This circuit is typical of those found in VCA controlled audio mixers.

In the early days of radio all program material was live and the audition channel served as the cue channel. Levels were set and program material was previewed using the audition channel. Today the audition channel records network feeds for later use while the program audio travels the program channels. The audition channel can monitor the audio quality and set levels of remote feeds using the studio monitor system.

Some consoles allow audio from the mixers to feed the program and audition channels simultaneously. This allows recording of on-air programming using the audition channel. In talk formats using fixed-time delay systems, the audition channel mixes the real-time program audio and feeds it to the delay. The output of the delay is brought back in to the console on the program channel, which then feeds the delayed program to the transmitter.

Television Audio Consoles

The requirements of TV audio are *compounded* because more live audio sources come into play (as opposed to radio, where most program material is recorded). In television installations additional monitoring requirements mean multiple output mixes. Television consoles include mono input modules for microphones, telephone hybrids and other mono sources. The inputs offer a gain control or switch to allow stepping between microphone and line-level input. These modules include a pan pot to allow left-right positioning of the apparent audio source. A mode switch may replace the pan pot, enabling selection of normal stereo, left channel only, right channel only, a mono mix or reversed channels. A cue channel feed plus a solo button permits stereo monitoring of a single audio source in the control room monitors. Audio sweetening requirements make equalization on each module a popular option.

The ability to create multiple audio feeds by the use of submaster mixing buses represents the major departure from radio consoles. The operator assigns mixer outputs to submaster buses and these submasters in turn create the master mix. This allows creation of several mix-minus feeds for special monitoring requirements.

Console Features and Options

A growing list of optional equipment available on audio consoles serve both radio and television audio mixing. These make them more user-friendly and provide expanded capabilities. Clocks and timers put timing functions in the immediate field of vision of the board operator. The timer resets to zero anytime that a new channel is selected so the operator will know how long a CD or cart has been running.

A mix-minus, program audio minus the caller's voice, feeds console to the telephone hybrid then down the line to the caller. If program audio, including the caller's voice, were fed to the phone hybrid a feedback path would exist. Figure 3.5-5 illustrates a block diagram of a mix-minus circuit for telephone hybrid. This example demonstrates that not all console inputs must route to the mix-minus bus. Notice that only Input

Studio Wiring Plan			
	Input Jack	Output Jack	
	Patch Panel #2	Patch Panel #2	
Program Output	Jacks 1 & 2	Jacks 25 & 26	AGC Amplifier Input
AGC Amplifier Out	Jacks 3 & 4	Jacks 27 & 28	Limiter Input
Limiter Output	Jacks 5 & 6	Jacks 29 & 30	STL Transmitter
Spare	Jacks 7 & 8	Jacks 31 & 32	Spare
Audition Output	Jacks 9 & 10	Jacks 33 & 34	Reel to Reel
Spare	Jacks 11 & 12	Jacks 35 & 36	Prod. Console In
Mono Output	Jacks 13 & 14	Jacks 37 & 38	Office Monitor
Mix-Minus Output	Jacks 15 & 16	Jacks 39 & 40	Telephone Hybrid
Spare	Jacks 17 & 18	Jacks 41 & 42	Spare
Prod. Console Out	Jacks 19 & 20	Jacks 43 & 44	Spare
Spare	Jacks 21 & 22	Jacks 45 & 46	Spare
Spare	Jacks 23 & 24	Jacks 47 & 48	Spare

Figure 3.5-3. This example of studio wiring of the console shows the versatility afforded by patch panels. Should the main console fail, patch cords can feed the production studio to the on-air processing and the transmitters.

Buying Equipment

Preparation of a shopping list and cost estimate for new equipment follows the design of the studio on paper. Choose the major components (console, cart players, professional CD players and furniture) early in the planning stage so that budget cuts will not compromise their quality. Allow for price increases, sales tax and shipping costs. Watch out for hidden handling or drop-shipment charges.

Each engineer should maintain a working relationship with a reputable broadcast equipment dealer. Absolute bottom dollar may not be the best deal. Consider the absence of real savings in paying $15.00 less for a CD player that fails to arrive in time to make the on-air date for the new studio. Experienced broadcast equipment salesmen offer their best deals and service to customers to whom they sell on a regular basis.

Fax the shopping list to the dealer at least a week before you need the cost estimate in order to allow the dealer time to research and work up a quote. Engineers who demand quotes on short notice seldom get serious attention by dealers. Competition in the broadcast supply business ensures that pricing between reputable dealers will vary by only a few percentage points.

If management insists on competitive bids limit this exercise to two bids. Time is more valuable than chasing nickels and dimes. If your regular dealer does not offer an item required for the project ask them to recommend a source. A salesperson will know all good suppliers and sometimes offer to get equipment not in the normal line for their best customers. This extra service can be worth a lot more than a few dollars when considering the big picture.

Audio Consoles

Centered in the radio studio, in front of the disc jockey sits the on-air audio console. An 8 to 12 mixer analog console typifies this unit. Smaller mixing consoles find their way into news editing rooms and production studios.

The number of mixing channels limits the number of audio events which can occur simultaneously or in rapid succession. The station format dictates its requirements. Although an operator-assisted easy listening or satellite based format may be able to use a four or five channel console; it would be out of the question for a fast-paced contemporary or rock program. These smaller consoles may not offer an audition bus or switchable inputs. The lack of multiple switched inputs requires one mixer for each audio source. External switchers can provide extra flexibility when required.

Wear and build-up of dirt plague console attenuators which depend on sliding a contact over a resistive element. The resistive element consists of either carbon, conductive plastic or metal film. Normal wear changes the element's resistance and build-up of worn-off carbon may cause erratic resistive changes in the contact between the slider and the element. Noise and uneven tracking between stereo channels results with age.

In years past noisy pots presented a major problem when the program audio ran directly through the potentiometer. Nowadays, however, most consoles employ voltage controlled amplifiers (VCAs) to avoid this problem. In modern consoles audio does not pass through the potentiometer. Instead, only a sample dc control voltage passes through it. This voltage controls the gain of an amplifier which carries the program audio. Figure 3.5-4 offers a schematic of a VCA control circuit.

VCA console design solves the problem of audible noise from defective attenuators, but the noisy attenuator will affect the accuracy of the control voltage. The audio passing through the VCA controlled circuit will become erratic and nonlinear as the noisy control moves through its range. Replacement of the defective attenuator prevents clumsy sounding cross-fades and stereo channel dropouts. This console design allows the use of cheaper rotary or straight-line (slider) attenuators. As with any amplifier, the VCA introduces some thermal noise and distortion. A good console has an overall distortion figure of 0.05% total harmonic distortion (THD) or less. The noise floor should be below −90 dBm.

Modular design consoles offer the engineer major advantages. Removing and swapping modules greatly simplifies troubleshooting. The layout of the console allows easy changes to keep up with station format undulations. The positioning of blank panels create dividers that neatly group sources together. Custom panels accommodate special functions such as reel-to-reel recorder control or telephone line selection. Extra space left in the mainframe offers expansion with additional mixers as the station's needs grow. These benefits justify the extra investment in a modular console.

For on-air use, a cue channel allows the operator to receive cue audio from remotes and networks and preview program material. Even if carts deliver all program material, a cue channel is a must. It serves as a valuable trouble-shooting tool.

Accepted engineering practice runs all line level inputs through patch panels on their way to console inputs. This allows rerouting special program audio and patching around any problems that may develop. The exception to the rule is microphone level audio. Directly wiring microphone outputs to the console input terminals remains the best option. Microphone input positions seldom change and the extra wiring through patch panel jacks invites noise problems. A possible exception is the television studio where the program must allow transfer from set to set, in which case adding a mic preamp to boost the signal to line level before patching may be appropriate.

Part of proper planning for a new console ensures that levels from all sources will be compatible with the input levels required by the console. If not properly matched, the operating positions of the potentiometers (pots) will be different for each mixer, making it difficult for the operator to run the board properly. The operator may open the pot only a fraction of a turn and drive the meters to the pin or may not get enough gain even with the control fully open. Either situation results in a poor audio mix with possible distortion and noise problems. Normal operation sets rotary attenuators at the two o'clock position and slider attenuators at a 70% position.

All studio sources should be adjusted to operate within their normal output range. Interface amplifiers (matching boxes) should be used to boost low level audio sources to the +4 dB levels required for most consoles. If the source signal overloads the console input an H-pad will drop it to gain a proper match.

An H-pad is a simple network of five resistors arranged in the shape of an "H" laid on its side. The resistors convert some of the audio energy into heat which drops the audio level by a predictable amount. At the same time, the H-pad maintains the impedance of the circuit. Pads of 10 dB, or greater, also correct impedance mismatches. Figure 3.5-1 illustrates the schematic of the resistive H-pad.

Pads are a must when directly connecting two audio transformers directly. If an output transformer directly feeds an input transformer, the output transformer sees a changing inductive load and frequency response and the system suffers. With an H-pad between them, the output transformer sees a linear resistive load. The H-pad can be mounted inside the equipment or at the studio punch blocks. Table 3.5-1 shows resistor values for commonly needed 600 Ω H-pads.

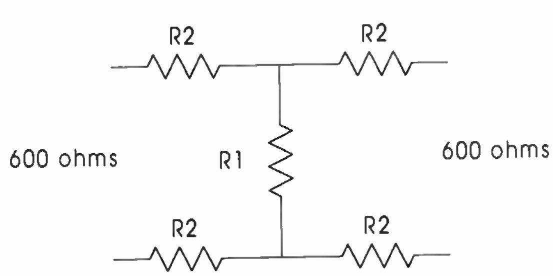


Figure 3.5-1. The H-pad resistive network reduces levels and matches impedances in balanced analog audio circuits.

Table 3.5-1
Typical H-Pad Values

Loss (in dB)	R1 (Ohms)	R2 (Ohms)
1	5100	18
3	1800	51
6	820	100
8	560	130
10	430	160
12	330	180
14	240	200
16	200	220
18	150	240
20	120	240
22	100	270
24	75	270
26	62	270
28	47	270
30	39	270
32	30	300
34	24	300
36	18	300
38	15	300
40	12	300

The worksheets in Figure 3.5-2 and Figure 3.5-3 offer an example of the starting point for designing a studio. Figure 3.5-2 illustrates wiring to the console inputs; Figure 3.5-3 designates the wiring path of the console outputs. Preparing customized worksheets for each studio saves false starts and wiring errors. Any computer spreadsheet program accommodates the task easily. Computer drafting programs offer another option for perfecting studio wiring prior to pulling cable.

The studio outlined in the worksheets is one of two in a station. A second studio serves as the production room. Patch panels route all audio. Outputs of both studios appear at the patch panel in the main studio as well as the production studio. If technical problems take the control room off-line the station can originate the program from the production console.

Studio Wiring Plan			
Audio Source	Input Jacks	Output Jacks	Console Inputs
	Patch Panel #1	Patch Panel #1	
Microphone #1			Mixer 1
Microphone #2			Mixer 2
CD #1	Jacks 1 & 2	Jacks 25 & 26	Mixer 3
CD #2	Jacks 3 & 4	Jacks 27 & 28	Mixer 4
Cart #1	Jacks 5 & 6	Jacks 29 & 30	Mixer 5
Cart #2	Jacks 7 & 8	Jacks 31 & 32	Mixer 6
Cart #3	Jacks 9 & 10	Jacks 33 & 34	Mixer 7
Satellite #1	Jacks 11 & 12	Jacks 35 & 36	Mixer 8
Satellite #2	Jacks 13 & 14	Jacks 37 & 38	Mixer 9
Reel to Reel	Jacks 15 & 16	Jacks 39 & 40	Mixer 10
Phone Hybrid	Jacks 17 & 18	Jacks 41 & 42	Mixer 11
EAS Receiver	Jacks 19 & 20	Jacks 43 & 44	Mixer 12
Production Studio	Jacks 21 & 22	Jacks 45 & 46	Mixer 13
Spare	Jacks 23 & 24	Jacks 47 & 48	Mixer 14

Figure 3.5-2. Careful planning of the console inputs prevents delays and rework during the installation process.

Digital audio equipment presents another level matching challenge. Headroom meters, not VU meters, monitor digital inputs. A headroom meter indicates how close the input signal comes to clipping in the analog to digital converter circuit. A 0 dBm signal fed to a digital recorder with 18 dB of headroom will appear on the headroom meter at − 18 dB. Headroom meters referencing a decibel value are identified as dBfs (*decibels* referenced to *full-scale*).

Analog circuits tolerate operation with the levels driven *in the red* above 0 dBm on their VU meters. Zero dBm is not the *clip point*. An analog circuit clips when the audio signal exceeds the voltage potential of the power supply. A sine wave then flattens on the peaks when the input signal reaches a level higher than the power supply voltage. This condition results in audio distortion.

Typically, analog audio equipment clips around + 24 dBu. If the manufacturer calibrated the VU meter to an output of + 4 dBu ("0" on the meters equals an output of + 4 dBu), the equipment is said to have 20 dB of headroom. Input audio reading an average value of 0 dB could contain audio peaks 20 dB higher without clipping and distorting as it passes through the equipment. It is easy to understand why analog equipment with 20 dB of headroom forgives trespasses into the *red* above 0 dB on the VU meter.

Digital equipment inputs demand closer attention. When a digital recorder, with a headroom meter, is driven above "0," the digital audio clips and the recording will contain non-repairable distortion, clicks or pops. Digital audio clips at the analog-to-digital converter (A/D). Digital clipping occurs when the analog input signal drives the A/D past its maximum output capability. Digital inputs should be operated as close to the "0" mark as possible, but do not clip! Good digital recording practice maintains peaks of − 6 dBfs.

Professional analog broadcast equipment features balanced inputs and outputs. A balanced audio output consists of two wires which carry the analog audio voltage to the next device. Neither of these wires connect to ground. Only the cable shield, which protects the audio from electrical noise and hum, is grounded. A balanced circuit can be run over properly shielded audio cables several hundred feet long.

When wiring balanced analog audio, use care to connect the high, or plus (+), terminals. Connect the low, or negative (−), only to similarly marked terminals. If you use 3-pin XLR connectors, pin #2 is always high (+). Pin #3 is always low (−). Pin #1 of an XLR connector is ground. When you use ¼ in. tip ring sleeve (TRS) plugs, the tip is high (+), the ring is low (−) and the sleeve is ground. Failure to follow the rules of polarity will result in out-of-phase audio. Stereo audio, wired out-of-phase, results in an audio dead spot centered in front of the speakers. Listeners with monaural radios hear only the difference between left channel and right channel if out-of-phase audio is fed to the transmitter.

An unbalanced audio circuit consists of a single conductor and grounded shield. Consumer electronic equipment uses unbalanced audio circuits easily identified by the single-pin, RCA phono plugs found on the connecting cables. An unbalanced circuit carries half of the power of a balanced circuit. It is more subject to hum and noise. Unbalanced circuits cannot support long runs of cable for this reason.

Bridging audio input circuits present no load to the source audio. Bridging inputs abandon the traditional 600 Ω, power-matching, input circuits found on older equipment. Without a load, no power transfer takes place. These bridging circuits simply transfer an audio voltage from output to input. A bridging input handles a wider variety of input sources than possible when everything terminated with a 600 Ω load. If an output requires a 600 Ω termination, a 620 Ω resistor tied across the input terminals provides a reasonable match.

In the old days of tube equipment, both inputs and outputs used transformers. Tubes could not drive 600 Ω loads directly; solid-state equipment could. Audio equipment manufacturers slandered the transformer with reports of poor performance as they removed them during the transition to solid-state design. Good transformers are expensive; manufacturers looked for ways to cut their costs and the audio transformer became a casualty.

Some applications still require the physical isolation that only a transformer can provide. Any time that audio equipment connects to a phone line, a transformer blocks the 48 V telco battery from entering the equipment. Unusually long audio cable runs operate best, with lower noise, when a transformer isolates the equipment on each end. The audio transformer provides best common mode rejection of electrical noise induced into long cable runs. Transformers still provide input termination on may high-end microphone preamplifiers. The transformer provides the required 150 Ω load for the microphone and isolates the preamp from the phantom voltage required to power condenser microphones. Common mode rejection of noise becomes even more important when dealing with the extremely low level output signals of dynamic microphones.

Step One: Where to Begin

Planning a new studio, or rebuilding an old one, begins with a layout on paper, or computer screen, of all of the required audio sources and feeds. The console inputs offer a good place to start. Working from a list of all possible sources the engineer assigns them priorities according to how often and how quickly the operator must put them on the air. This determines the number of mixing channels needed and how many switched inputs each mixer requires.

All frequently used audio sources should be assigned to individual console mixers. Keeping input switching and patching required of the operator to a minimum avoids errors and dead-air. A console with two or three more inputs and mixers than absolutely necessary provides inexpensive insurance against obsolescence and frequent studio rewiring jobs.

events around, control audio source equipment, adjust levels and even read copy and tags directly from the screen.

Talk formats require a studio for the show host and on-air guests. A large round table with the microphone booms mounted in the center allows the host and any guests working room plus affords eye contact. The show producer operates the mixing console and telephone hybrids in the control room. If not removed to an adjacent room, the show's call screener also shares the space in the control room.

Large market budgets afford lots of resources and first-class equipment. A visit to a small market talk show may reveal the show host running all of the control room equipment while talking with callers. The station's telephone receptionist works frantically screening calls and shuttling caller names on yellow Post-It-Notes® to the show host.

The Production Studio

A separate studio provides an area for commercial production work and transfer of music to cart or to a digital audio storage system. The production studio provides access to CD players, a digital audio workstation, cart recorders, equalization, patch panels, reel-to-reel recorders and possibly turntables. The production studio provides more flexibility and handles a wider variety of audio mediums than the on-air studio.

Four, eight and sixteen-track consoles and multi-track recorders found in major market production facilities often intimidate the typical disk jockey in segueing music and commercials. These consoles feature sub-master mixing busses allowing the mix down of multiple tracks of music, voice tracks and sound effects producing complex spots and promos. Special effects and equalization, not needed in the main studios, remain standard fare in production.

Production directors prefer digital workstations because they allow editing without splicing tape. Digital workstations bring to audio production the speed and versatility that word-processing brought to typing. Workstations allow editing of individual tracks, a feat impossible with multi-track reel-to-reel tape editing.

Since all material produced in the studio will eventually be played on the air, the quality of the equipment should be equal to, if not better than, that used in the main studio.

Typical Television Studio Layout

Audio for television stations can be challenging because of the need for many types of audio mixes. Multi-channel television sound (MTS) requires a stereo program feed to the transmitter, on-camera talent needs a monitor mix, programs with a studio audience require a mono PA mix, talk and news programs must have a mix-minus for telephone hybrids and possibly a mix in a different language for a second audio program (SAP) channel.

The ATSC DTV Standard will make life even more interesting. As the motion picture format 5.1 channel comes to television, the audio engineer must deal with left channel, center channel, right channel, two surround channels and a low-frequency effects (LFE) channel. The first five audio channels offer full 20 kHz bandwidth. The sub-woofer LFE channel provides response from 3 Hz–125 Hz. All of this arrives as digitally compressed audio in the Dolby AC-3 format.

Audio monitoring of the preview channel is accomplished through the mixer solo function. If a remote truck is involved, an interruptible foldback (IFB) system requires a program audio feed plus audio cues to the talent at the truck. The console to handle all this requires a design specifically for these complex tasks. An audio routing switcher tied to the video switcher assumes some of the work load.

STUDIO PLANNING

Analog audio signals consist of complex mixtures of alternating currents of different frequencies at different powers. In broadcasting, the decibel serves as one unit for measuring audio power. Power is measured with VU (volume units) meters, marked in decibels (dB), on broadcast equipment. The decibel is one tenth of the logarithm of the ratio of two powers. Volume units are referenced to the 1 mW of power produced by a 1 kHz audio tone across a 600 Ω load. This combination of audio power (1 mW) and load value (600 Ω) creates 0.775 V of audio and is represented by the "0" on the VU meter.

The designation "dBm" identifies all readings referenced above or below this power level. A signal of −3 dBm contains half of the energy of a signal measured at 0 dBm. A signal of +3 dBm is twice as loud as the 0 dBm signal. A signal of +10 dBm is 10 times as powerful.

Decibel values identified as "dBu," (decibels-unterminated) indicate an audio voltage of 0.775 V across an open circuit. Decibels measured in "dBv," (decibels-volts) are also measured across an open circuit, but are referenced to 1.0 V. Since the circuit lacks the traditional 600 Ω load, these are simple measurements of an audio voltage, not power. Modern analog equipment no longer uses 600 Ω, impedance matching, circuits; most references to decibels now carry the "dBu" or "dBv" notation.

Occasionally the 0 dBm output level on a console drives the input meter on a recorder to +4 dBm. Zero dBm still equals 0 dBm, but equipment manufacturers often calibrate their equipment to provide an output level of +4 dBm, or +8 dBm, when the output VU meter indicates 0 dBm. Engineers always standardize operating levels of all equipment in the station. This allows patching any output into any input without operating level problems.

Signal levels of 0 dBm, +4 dBm and +8 dBm are all valid operating levels. Each station may use a different level, but an engineer will calibrate all of the equipment in the station to one of these three standard operating levels. Someone bringing in equipment from outside the facility may find it necessary to recalibrate to the signal level used throughout the facility.

3.5
STUDIO AUDIO EQUIPMENT

WALT LOWERY
HARRIS CORP., QUINCY, IL

INTRODUCTION

The quality of a radio station's only distributed product, its sound, is determined by the events taking place in the studio. The best source material and the best air talent will produce only marginal results when burdened by inadequate equipment. Even in television, audio can no longer be treated as a secondary technology. State of the art equipment, properly used in the studio, equates to ratings and translates to revenue for the station.

Broadcast studios now take advantage of digital audio source material, digital audio processing equipment and digital consoles. As music CDs replaced vinyl discs as the standard source material for radio broadcasters over an alarmingly short period of time, digital broadcast equipment is poised to eclipse analog studio systems. We have the potential to reduce the radio studio to a touch-screen computer operation, but the trend in the industry remains loyal to familiar function and feel. Broadcasters prefer their digital equipment to emulate the friendly analog devices which they have used for decades.

TYPICAL RADIO STUDIO LAYOUTS

Many stations in active markets have suffered through changes in formats, managers and engineers leaving a crazy quilt of analog and digital equipment. The results have handicapped the station's ability to survive in a competitive market. Programming needs, available space and creativity of former engineers often dictated the design of a studio. Design du jour, compounded with galloping changes in technology, leaves many studios ripe for redesign and rebuilding.

The station's current format should dictate the main design parameters. The operational concepts will be much different for a music format than for a news/talk operation. It is possible to catalog radio station studios under a few general categories.

Music Formats

Studios built for music formats remain the most common, particularly in smaller markets. The basic configuration consists of an audio console, two or more CD players and multiple cart players arranged in a "U" or "L" shaped desk. Other necessary equipment might include audio routing, transmitter remote control and telephone interface equipment. Careful design places all equipment within arm's reach of the operator. This configuration handles the rather simple programming needs nicely. The announcer on duty mixes and switches audio, reads announcements, plays music from CD players, plays commercials from analog or digital cart players and switches news and network programming from a satellite receiver.

Some stations transfer music from CD to digital carts, making the studio an all-cart operation. In this situation the studio contains four to six cart playback decks and no CD players. Having all program material on cart makes the operation more simple. The operator needs only to talk, slam plastic discs into slots and push buttons.

Going one step further, some stations retrieve their program material from digital audio storage systems. Commercials and music are stored digitally on a hard drive, retrieved when required, converted back to analog audio and fed to the console. This eliminates the physical handling of any audio storage medium. The operator controls each audio event with a keyboard and mouse.

Music formats require positions for one combo operator and possibly an announce booth for news. With the popularity of the *Morning Zoo* format, many large market stations designed their studios for two or three on-air personalities. One member of the team operates the console, another juggles phone calls, pulling music and spots and the third handles news. The studio layout varies with the duties handled by each member of the team.

News/Talk Formats

All-news stations serve up information from live in-studio talent and short duration feeds from a large number of sources. Here the board operator functions as an engineer rather than on-air talent. Small studios orbit the central control room allowing eye contact between the board operator and on-air talent. The console provides more of a switching function than audio mixing. Live news or talk programs dictate a larger number of console inputs than the typical disk jockey operation. Digital audio storage systems help this format flow smoothly.

Technology now allows replacement of the audio console with a computerized audio switching system. Hard disk audio storage systems utilize multiple computers linked to a central network server through a local area network (LAN). Each studio retains access to all stored program material. The operator sees the log displayed on the screen. The operator can shift

This random access memory device has a capacity measured in hundreds of gigabytes. The data is stored as three-dimensional optical holograms, with data being written and retrieved as two dimensional patterns of laser light. A light sensitive crystal, (Holostore) serves as the medium for PVHS.

This form of crystal memory is non-volatile, removable, and achieves data transfer rates of 1 TB (terabyte) per second. Today's fastest magnetic disk would take five hours to resolve the same amount of data. It has been reported that a removable PVHS media of some 10 × 10 × .5 cm could store 100 GB—about a month's worth of continuous stereo digital audio. Over the coming decade the emergence of static optical storage may reduce moving magnetic media applications to antiquity.

BIBLIOGRAPHY

Alten, Stanley R., *Audio in Media*, Fourth Edition, Wadsworth Publishing Company, Belmont, California, 1994.

Audio Engineering Society, "AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Linearly Represented Digital Audio Data," *Journal of the Audio Engineering Society*, vol. 33, no. 12, December, 1985.

Carasso, M.G., Peck, J.B.H., Sinjou, J.P., "The Compact Disc Digital Audio System," *Philips Technical Review*, vol. 40, no. 6, 1982.

EBU (European Broadcasting Union). "Specification of the Digital Audio Interface." EBU Doc. Tech., 3250.

IEC (International Electrotechnical Commission). "Draft Standard for a Digital Audio Interface." IEC Report TC 84/WG11, November, 1986.

Jorgensen, Finn, *The Complete Handbook of Magnetic Recording*, 4th Edition, McGraw-Hill, 1996.

Mee, Dennis C. & Daniel, Eric D., editors, *Magnetic Recording Technology*, Second Edition, McGraw-Hill, 1996.

Pohlmann, Ken C., *Principles of Digital Audio*, Third Edition, McGraw-Hill, 1995.

Pohlmann, Ken C. *The Compact Disc*, Second Edition. AR Editions, Madison, WI, 1992.

Pohlmann, Ken C., *Principles of Digital Audio*, Second Edition. Howard W. Sams and Co., Carmel, IN, 1989.

Pohlmann, Ken C., editor, *Advanced Digital Audio*. Howard W. Sams and Co., Carmel, IN, 1991.

Watkinson, John, *RDAT*, Focal Press, Oxford, 1991.

Additional Sources of Information

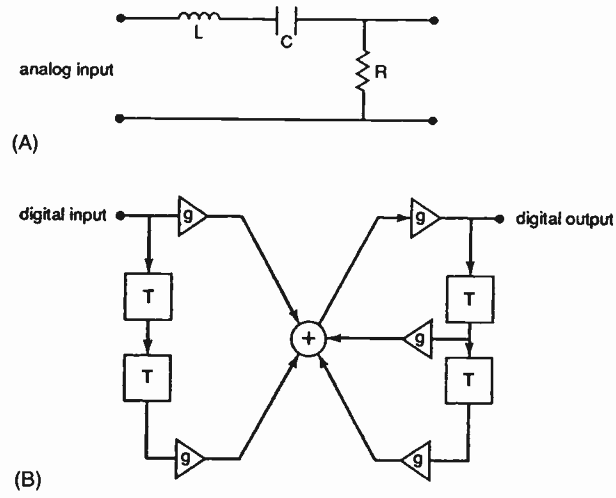
Tremaine, Howard M., *Audio Cyclopedia*, Howard W. Sams, Inc., now out of print but still available via specialty bookstores.

and expansion, crossovers, reverberation, ambience processing, time alignment, acoustic noise cancellation, mixing and editing and acoustic analysis. Some DSP functions are embedded within other applications; for example, the error correction systems and oversampling filters found in CD players are examples of DSP. In other applications the user has control over the DSP functions.

Digital processing is more precise, repeatable and can perform operations that are impossible with analog techniques. Noise and distortion can be much lower with DSP thus audio fidelity is much higher. In addition, whereas analog circuits age, lose calibration and are susceptible to damage in harsh environments, DSP circuits do not age, cannot lose calibration and are much more robust. However DSP technology is an expensive technology to develop. Hardware engineers must design the circuit or employ a DSP chip, and software engineers must write appropriate programs. Special concerns must be addressed when writing the code needed to process the signal. For example, if a number is simply truncated without regard to its value, a significant error could occur, and the error would be compounded as many calculations take place, each using truncated results. The resulting numerical error would be manifested as distortion in the output signal. Thus all computations on the audio signal must be highly accurate. This requires long wordlengths; DSP chips employ digital words that are 32 bits in length, or longer.

In addition, even simple DSP operations may require several intermediate calculations and complex operations may require hundreds of operations. To accomplish this, the hardware must execute the steps very quickly. Because all computation must be accomplished in real time, that is, within the span of one sample period, the processing speed of the system is crucial. A DSP chip must often process 50–100 million instructions per second. This allows it to run complete software programs on every audio sample as it passes through the chip.

DSP products are more complicated than similar analog circuits, but DSP possesses an inherent advantage over analog technology—it is programmable. Using software, many complicated functions can be performed entirely with coded instructions. Figure 3.4-21(a) shows a band-pass filter using conventional analog components. Figure 3.4-21(b) shows the same filter, represented as a DSP circuit. It employs the three basic DSP operators of delay, addition and multiplication. However this DSP circuit may be realized in software terms. Figure 3.4-21(c) shows an example of the computer code (Motorola DSP56001) needed to perform band-pass filtering with a DSP chip. There are many advantages to this software implementation. Whereas hardware circuits would require new hardware components and new circuit design to change their processing tasks, the software implementation could be changed by altering parameters in the code. Moreover the program could be written so different parameters could be employed based on user control.




```
MPY    X0,Y1,A    X: (R0) +, X0    Y: (R4) +, Y0
MAC    X0,Y0,A    X: (R0) +, X0    Y: (R5) +, Y0
MAC    X0,Y0,A    X: (R0) +, Y0    Y: (R5) +, Y0
MAC    X0,Y0,A    X: (R0) +, X0    Y1,Y: (R4) +
MOVE              A,X1             A,Y: (R5)
```

(C)

Figure 3.4-21. A band-pass filter represented by an analog circuit (a), digital signal processing circuit (b), and digital signal processing instructions (c). (Courtesy Motorola.)

As noted, DSP can be used in lieu of most conventional analog processing circuits. The advantages of DSP are particularly apparent when various applications such as recording, mixing, equalization and editing are combined in a workstation. For example, a personal computer, combined with a DSP hardware card, hard disk drive, appropriate software and a DAT or CD recorder forms a complete post production system. Such a system allows comprehensive signal manipulation including ability to cut, paste, copy, replace, reverse, trim, invert, fade in, fade out, smooth, loop, mix, change gain and pitch, crossfade and equalize. The integrated nature of such a workstation, its low cost, and high processing fidelity make it clearly superior to analog techniques.

ADVANCED RESEARCH AND DEVELOPMENT

A continuing limitation of all common analog or digital audio systems is the requirement for moving media. The sophistication of high fidelity rotation and pickup devices inherently necessitate certain minimum sizes and production costs. Although some small solid-state storage media exist the volume of data and high price of memory prohibits today's technology being applied to professional audio recording systems.

Radical static storage approaches are being researched and at least one, photorefractive volume holographic storage (PVHS), shows impressive promise.

and odd data blocks, one for each head; half of each channel's samples are recorded by each head. All of the data are encoded with a doubly-encoded Reed-Solomon error correction code. The error correction system can correct any dropout error up to 2.6 mm in diameter, or a stripe 0.3 mm high. Dropouts up to 8.8 mm long and 1.0 mm high can be concealed with interpolation.

Serial Interfacing

Most professional digital audio devices employ an output protocol using the joint Audio Engineering Society (AES) and European Broadcasting Union (EBU) serial transmission format for digital audio data. It is known as the AES/EBU or AES3 format and is specified in the ANSI S4.40-1992 Standard. Manufacturers of consumer electronics have adopted a derivative transmission format that has been standardized by the International Electrotechnical Commission (IEC) and is commonly referred to as the SPDIF. It is specified in the IEC Report TC 84/WG11 and IEC Publication 958.

The AES/EBU digital audio format transmits and receives left and right channel data using one digital cable. The transmission rate corresponds exactly to the source sampling frequency. One frame consists of two subframes, labeled A (left channel) and B (right channel), each with 32 bits. Each subframe contains data for one audio channel. The first 4 bits are used for synchronization and identifying preambles. The next 24 bits carry audio data, with the MSB transmitted last; 16 bit audio data leaves 4 bits at zero; the first 4 bits in the field are set aside for auxiliary audio or other data, as shown in Figure 3.4-20.

The last 4 bits form a control field with the V, U, C, and P bits. The validity (V) bit indicates if the previous audio sample is error free. The user (U) bit can be used to form a block of user data associated with the audio channel. The channel (C) status bit is used to form a data block; for each channel, one block is formed from the channel status bit contained in 192 successive frames. The parity (P) bit is used to provide even parity for each subframe. The AES/EBU Standard specifies that data are transmitted over twisted pair conductors, with 3-pin XLR connectors. There is a supplemental document, called AES-3id–1995, that specifies the method for transmission of AES formatted data by unbalanced coaxial cable.

In the IEC or SPDIF serial format used in consumer digital audio equipment the first bit of the channel status byte is set to "0" to signify a consumer interface, and a different channel status specification is used. In addition, when interfacing CD data, provision is made to transmit the subcode data in the user bit channel. The consumer format also contains provision for SCMS in the channel status bits. In particular, bit 2 is used to flag copyrighted material, and bit 15 distinguishes between original and copied material. In the IEC or SPDIF format, video coaxial cable with phono plugs or fiber optic cables can be used to convey data.

Digital interconnection formats such as AES3 and SPDIF allow audio data to be transferred from one device to another without any generation loss whatsoever. For example, audio data from a DAT recording could be conveyed to a hard disk editing system, processed there, and returned to DAT without transmission error. However it is important to only connect AES/EBU outputs to AES/EBU inputs, and IEC/SPDIF outputs to IEC/SPDIF inputs. AES/EBU and IEC/SPDIF interfaces should not be interconnected. Transmitted data could be invalid, or lead to improper machine operation

Digital Signal Processing

Digital signal processing (DSP) has improved the performance of many existing audio functions such as equalization and dynamic range compression, and permits new functions such as ambience processing, dynamic noise cancellation and time alignment. DSP is a technology used to analyze, manipulate or generate signals in the digital domain. It uses the same principles as any digitization system however instead of a storage medium such as CD or DAT, it is a processing method.

DSP Applications and Design

DSP employs technology similar to that used in computers and microprocessor systems however there is an important distinction. A regular computer processes data, whereas a DSP system processes signals. It is accurate to say that an audio DSP system is in reality a computer dedicated to the processing of audio signals.

Some audio functions DSP can perform include: error correction, multiplexing, sample rate conversion, speech and music synthesis, data compression, filtering, adaptive equalization, dynamic range compression

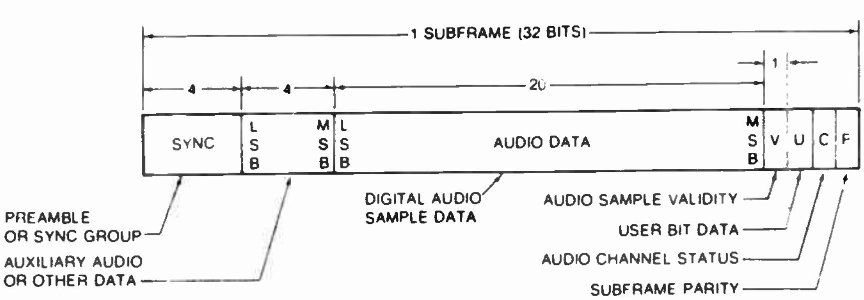


Figure 3.4-20. AES/EBU subframe format. (From Pohlmann, *Principles of Digital Audio.*)

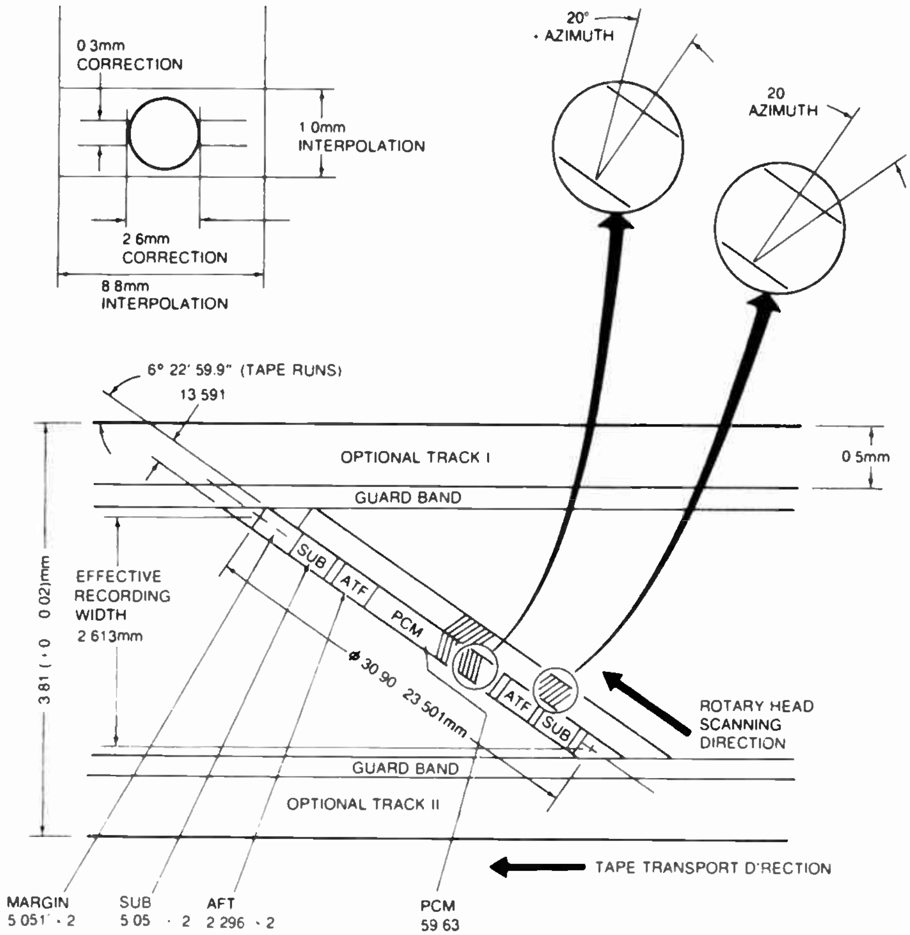


Figure 3.4-19. DAT track configuration. (From Pohlmann, *Principles of Digital Audio.*)

track types, sometimes referred to as A and B, with differing azimuth angles between successively recorded tracks. This ±20° azimuth angle means that the A head will read an adjacent B track at an attenuated level due to phase cancellation. This reduces crosstalk between adjacent tracks, eliminates the need for a guardband between tracks and promotes high density recording. Erasure is accomplished by overwriting new data to tape such that successive tracks partially write over previous tracks. Thus the head gaps (20.4 microns) are approximately 50% wider than the tracks (13.59 microns) recorded to tape.

The length of each track is 23.501 millimeters. Each bit of data occupies 0.67 microns, with an overall recording data density of 114 Mb per square inch. With a sampling rate of 48 kHz and 16 bit quantization, the audio data rate for two channels is 1.536 Mbps. However, error correction encoding adds extra information amounting to about 60% of the original, increasing the data rate to about 2.46 Mbps. Subcode raises the overall data rate to 2.77 Mbps.

The primary types of data recorded on each track are PCM audio, subcode and automatic track finding (ATF) patterns. Each data (or sync) block contains a sync byte, ID code byte, block address code byte, parity byte and 32 data bytes. In total, there are 288 bits per data block; following 8/10 modulation, this is increased to 360 channel bits. Four 8 bit bytes are used for sync and addressing. The ID code contains information on pre-emphasis, sampling frequency, quantization level, tape speed, copy-inhibit flag, channel number, etc. Subcode data are used primarily for program timing and selection numbering. The subcode capacity is 273.1 kbps. The parity byte is the exclusive or sum of the ID and block address bytes, and is used to error correct them.

Since the tape is always in contact with the rotating heads during record, playback and search modes, tape wear necessitates sophisticated error correction. DAT is thus designed to correct random and burst errors. Random errors are caused by crosstalk from an adjacent track, traces of an imperfectly erased signal or mechanical instability. Burst errors occur from dropouts caused by dust, scratches on the tape or by head clogging with dirt.

To facilitate error correction, each data track is split into halves, between left and right channels. In addition, data for each channel is interleaved into even

ITEM \ MODE	DAT (REC/PB MODE)				PRERECORDED TAPE(PB ONLY)	
	STANDARD	OPTION 1	OPTION 2	OPTION 3	NORMAL TRACK	WIDE TRACK
CHANNEL NUMBER [CH]	2	2	2	42	2	2
SAMPLING FREQUENCY [kHz]	48	32	32	32	44.1	
QUANTIZATION BIT NUMBER [BIT]	16 (LINEAR)	16 (LINEAR)	12 (NONLINEAR)	12 (NONLINEAR)	16 (LINEAR)	16 (LINEAR)
LINEAR RECORDING DENSITY [kBPI]	61.0	61.0			61.0	61.1
SURFACE RECORDING DENSITY [MBPI²]	114	114			114	76
TRANSMISSION RATE [MBPS]	2.46	2.46	1.23	2.46	2.46	
SUBCODE CAPACITY [KBPS]	273.1	273.1	136.5	273.1	273.1	
MODULATION SYSTEM	8–10 MODULATION					
CORRECTION SYSTEM	DOUBLE REED-SOLOMON CODE					
TRACKING SYSTEM	AREA SHARING ATF					
CASSETTE SIZE [mm]	73 x 54 x 10.5					
RECORDING TIME [MIN]	120	120	240	120	120	80
TAPE WIDTH [mm]	3.81					
TAPE TYPE	METAL POWER					OXIDE TAPE
TAPE THICKNESS [μm]	13 ± 1 μ					
TAPE SPEED [mm/s]	8.15	8.15	4.075	8.15	8.15	12.225
TRACK PITCH [μm]	13.591				13.591	20.41
TRACK ANGLE	6°22'59.5"					6°23'29.4"
STANDARD DRUM SPECIFICATIONS	ø30 90° WRAP					
DRUM ROTATIONS [rpm]	2000		1000	2000	2000	
RELATIVE SPEED [m/s]	3.133		1.567	3.133	3.133	3.129
HEAD AZIMUTH ANGLE	±20°					

Figure 3.4-18. DAT standard specifications. (From Pohlmann, *Principles of Digital Audio*.)

the helical scan track along with the audio signal, it is treated independently and can be rewritten without altering the audio program, and entered either during recording or playback. With the ID codes entered into the subcode area, desired points on the tape such as the beginning of selections can be searched for at high speed by detecting each ID code. During playback, if the skip ID is marked, playback is skipped to the point at which the next start ID is marked, and playback begins again.

In the DAT format, the recorded area is distinguished from a blank section of tape with no recorded signal, even if the recorded area does not contain an audio signal. Unlike blank areas, the track format is always encoded on the tape even if no signal is present. If these sections are mixed on a tape, search operations may be slowed. Hence, blank sections should be avoided. A consumer DAT deck with an interface meeting the specifications of the Sony Philips digital interface format (SPDIF) will identify when data have been recorded with a copy inhibit Serial Copy Management System (SCMS) flag in the subcode (ID6 in the main ID in the main data area) and will not digitally copy that recording. In other words, SCMS permits first generation digital copying, but not second generation copying. Analog copying is not inhibited.

DAT Recorder Design

From a hardware point of view, a DAT recorder utilizes many of the same elements as a CD-R recorder: A/D and D/A converters, modulators and demodulators, error correction encoding and decoding. Audio input is received in digital form, or is converted to digital by an A/D converter. Error correction code is added and interleaving is performed. As with any helical scan system, time compression must be used to separate the continuous input analog signal into segments prior to recording, then rejoin them upon playback with time expansion to form a continuous audio output signal. Subcode information is added to the bit stream, and it undergoes eight-to-ten (8/10) modulation. This signal is recorded via a recording amplifier and rotary transformer.

In the playback process the rotary head generates the record waveform. Track finding signals are derived from the tape and used to automatically adjust tracking. Eight-to-ten demodulation takes place and subcode data are separated and used for operator and servo control. A memory permits de-interleaving as well as time expansion and elimination of wow and flutter. Error correction is accomplished in the context of de- interleaving. Finally, the audio signal is output as a digital signal, or through D/A converters as an analog signal.

The DAT rotary head permits slow linear tape speed while achieving high bandwidth. Each track is discontinuously recorded as the tape runs past the tilted head drum spinning rapidly in the same direction as tape travel. The result are diagonal tracks at an angle of slightly more than 6° from the tape edge, as shown in Figure 3.4-19. Despite the slow linear tape speed of 8.15 mmps (1/4 in. per second), a high relative tape-to-head speed of about 3 msec (120 in. per second) is obtained. A DAT rotating drum (typically 30 mm in diameter) rotates at 2,000 rpm, typically has two heads placed 180° apart, and a tape wrap of only 90°. Four head designs provide direct read after write, so the recorded signal can be monitored.

Azimuth recording (*or guard-bandless recording*), is used in which the drum's two heads are angled differently with respect to the tape; this creates two these discs is vulnerable to sunlight thus discs should not be exposed to bright sun over a long period.

The CD-R format is defined in the *Orange Book Standard* authored by Philips and Sony. In CD recorders adhering to the *Orange Book I Standard*, a disc must be recorded in one pass—start-stop recording is not permitted. In recorders adhering to the *Orange Book II Standard*, recording may be stopped and started. In many players, tracks may be recorded at different times and replayed but because the disc lacks the final TOC, it can be played only on a CD-R recorder. When the entire disc is recorded the interim TOC data are transferred to a final TOC, and the disc may be played in any CD audio player. The program memory area (PMA) located at the inner portion of the disc contains the interim TOC record of the recorded tracks. In addition, discs contain a power calibration area (PCA); this allows recorders to automatically make test recordings to determine optimum laser power for recording. Some recorders exceed the *Orange Book II Standard*; they generate an interim TOC that allows partially recorded discs to be played on playback-only CD players.

CD-R recorders are useful because they eliminate the need to create an edited master tape prior to CD recording. If a passage is not wanted, it can be marked prior to writing the final TOC so that the recorder will not play it back. For example, dead air during a live performance can be marked so it is deleted whenever the disc is played back. The data physically continues to exist on the disc, however.

Erasable CD-E Format

CD systems that provide for both recording and erasing are known as *CD-E systems*. Erasable optical systems permit data to be written, read, erased and written again. Several recordable/erasable optical media have been introduced, most notably, magnetooptical (MO) media. Magnetooptical recording technology combines magnetic recording and laser optics, utilizing the record/erase benefits of magnetic materials with the high density and contactless pickup of optical materials.

With magnetooptics, a magnetic field is used to record data, but the applied magnetic field is much weaker than conventional recording fields. It is not strong enough to orient the magnetic particles. However the coercivity of the particles sharply decreases as they are heated to their Curie temperature. A laser beam focused through an objective lens heats a spot of magnetic material and only the particles in that spot are affected by the magnetic field from the recording coil, as shown in Figure 3.4-17(a). After the laser pulse is withdrawn, the temperature decreases and the orientation of the magnetic layer records the data. In this way, the laser beam creates a small recorded spot thus increasing recording density.

The Kerr effect may be used to read data; it describes the slight rotation of the plane of polarization of polarized light as it reflects from a magnetized material. The rotation of the plane of polarization of light re-

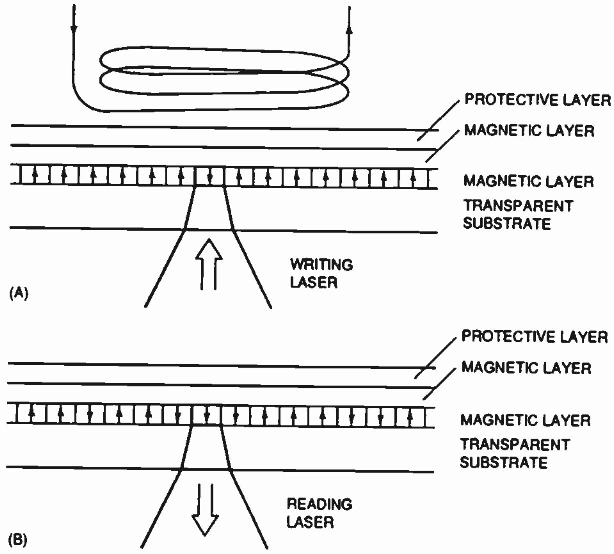


Figure 3.4-17. Magnetooptical recording (a) and playback (b).

flected from the reverse oriented regions differs from that reflected from unreversed regions, as shown in Figure 3.4-17(b). To read the disc, a low powered laser is focused on the data surface, and the angle of rotation of reflected light is monitored thus recovering data from the laser light. To erase data a magnetic field is applied to the disc, along with the laser heating spot. Tests indicate that MO media could be erased/recorded over 10 million times and would retain data for over 10 years.

Digital Audio Tape Format

The rotary-head digital audio tape (R-DAT or DAT) format was originally designed as a consumer medium to replace the analog cassette. However the format has found wider application as a low cost professional digital recording system.

Format Specifications

The DAT format supports four record/playback modes and two playback-only modes. The standard record/playback, and both playback-only modes, *wide* and *normal*, are implemented on every DAT recorder. The standard mode offers 16 bit linear quantization and 48 kHz sampling rate. Both playback-only modes use a 44.1 kHz sampling rate, for user- and prerecorded tapes. Three other record/playback modes, called Options 1, 2, and 3, all use 32 kHz sampling rates. Option 1 provides 2 hour recording time with 16 bit linear quantization. Option 2 provides 4 hours of recording time with 12 bit nonlinear quantization. Option 3 provides 4 channel recording and playback, also using 12 bit nonlinear quantization. These specifications are summarized in Figure 3.4-18.

The user can write and erase nonaudio information into the subcode area: start ID indicating the beginning of a selection, skip ID to skip over a selection, and program number indicating selection order. This subcode data permits rapid search and other functions. Although subcode data are recorded onto the tape in time constants of 15 and 50 μ sec. Upon playback, deemphasis is automatically carried out, resulting in an improvement in S/N.

Recordable CD-R Format

With a CD-R (or CD-WO) write-once optical disc recorder, the user may record data until the disc capacity is filled. Recorded CD-R discs are playable on conventional CD players. A block diagram of a CD-R recorder is shown in Figure 3.4-16. An encoder circuit accepts an input PCM signal and performs CIRC error correction encoding, EFM modulating, and other coding and directs the data stream to the recorder. The recorder accepts audio data and records up to 74 minutes in real time. In addition to audio data, a complete subcode table is written in the disc TOC, and appropriate flags are placed across the playing surface.

Write-once media is manufactured similarly to conventional playback-only discs. As with regular CDs, they employ a substrate, reflective layer and protective top layer. Sandwiched between the substrate and reflective layer, however, is a recording layer comprised of an organic dye. Together with the reflective layer it provides a typical in-groove reflectivity of 70% or more. Unlike playback-only CDs, a pregrooved spiral track is used to guide the recording laser along the spiral track; this greatly simplifies recorder hardware design and ensures disc compatibility. Shelf life of the media is said to be 10 years or more at 25° centigrade and 65% relative humidity. However the dye used in

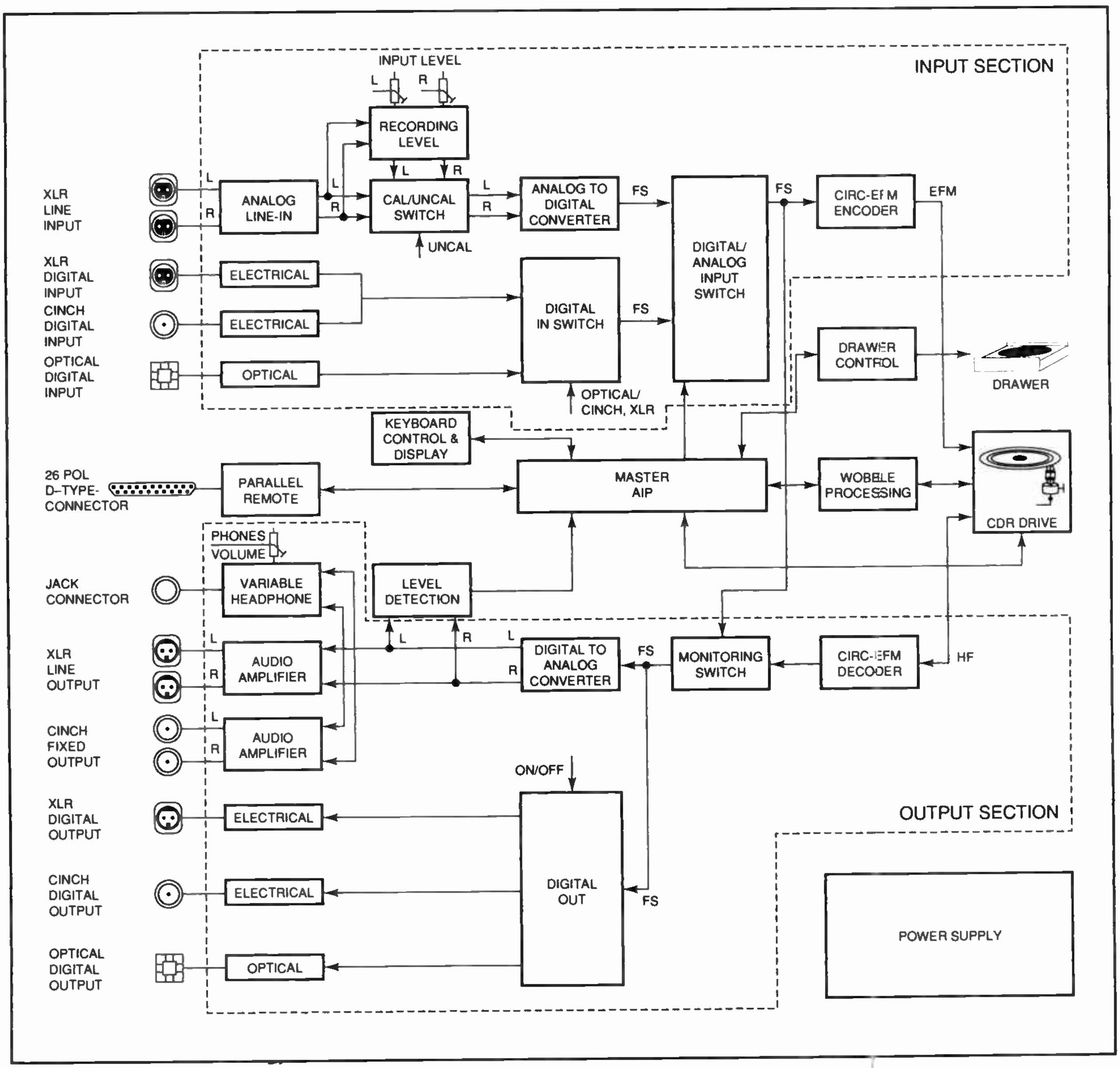


Figure 3.4-16. Block diagram of a CD-R recorder. (From Pohlmann, *The Compact Disc.*)

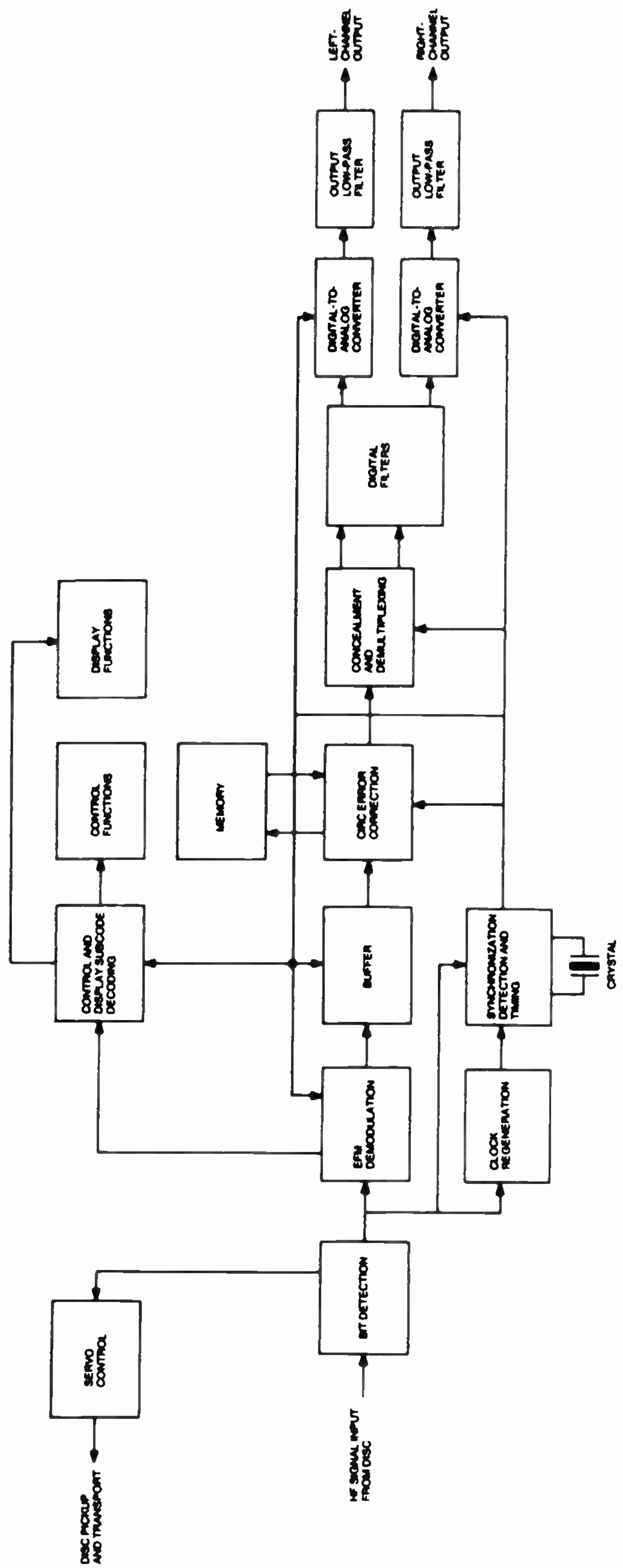


Figure 3.4-15. Block diagram of a CD player with digital filtering. (From Pohlmann, *The Compact Disc.*)

darker rings, and spot dimensions are quoted as half-power levels.

When viewed from the laser's perspective, the pits appear as bumps with height between 0.11–0.13 mm. This dimension is slightly less than the laser beam's wavelength in polycarbonate of 500 nm. The height of the bumps is thus approximately 1/4 of the laser's wavelength in the substrate. The reflective flat surface of a CD is called *land*. Light striking land travels a distance one-half wavelength longer than light striking a bump, as shown in Figure 3.4-14. This creates an out-of-phase condition between the part of the beam reflected from the bump, and the part reflected from the surrounding land. The beam thus undergoes destructive interference, resulting in cancellation. Optically, if the CD pit surface is considered as a two-dimensional reflective grating, the focused laser beam diffracts into higher orders, resulting in interference. The disc surface data thus modulates the intensity of the reflected light beam. In this way the data physically encoded on the disc are recovered by the laser.

Data Decoding

A CD player's data path, shown in Figure 3.4-15, directs the modulated light from the pickup through a series of processing circuits, ultimately yielding a stereo analog signal. Data decoding follows a procedure which essentially duplicates, in reverse order, the encoding process. The pickup's photodiode array and its processing circuits output EFM data as a high frequency signal. The first data to be extracted from the signal are synchronization words. This information is used to synchronize the 33 symbols of channel information in each frame, and a synchronization pulse is generated to aid in locating the zero crossing of the EFM signal.

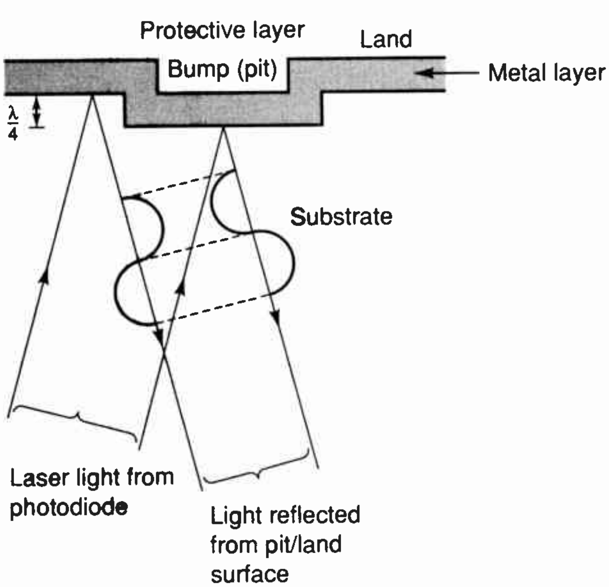


Figure 3.4-14. A pit causes cancellation through destructive interference.

The EFM signal is demodulated so that 17 bit EFM words again become 8 bits. A memory is used to buffer the effect of disc rotational wow and flutter. Following EFM demodulation, data are sent to a CIRC decoder for de-interleaving, and error detection and correction. The CIRC decoder accepts one frame of thirty two 8 bit symbols: 24 audio symbols and 8 parity symbols. One frame of twenty four 8 bit symbols are output. Parity from two Reed-Solomon decoders is utilized. The first error correction decoder corrects random errors and detects burst errors and flags them. The second decoder primarily corrects burst errors, as well as random errors that the first decoder was unable to correct. Error concealment algorithms employing interpolation and muting circuits follow CIRC decoding.

In most cases, the digital audio data are converted to a stereo analog signal. This reconstruction process requires one or two D/A converters, and low-pass filters to suppress high frequency image components. Rather than use an analog brickwall filter after the signal has been converted to analog form, the digitized signal is processed before D/A conversion using an oversampling digital filter. An oversampling filter uses samples from the disc as input, then computes interpolation samples, digitally implementing the response of an analog filter.

A *finite impulse response* (FIR) transversal filter is used in most CD players. Resampling is used to increase the sample rate; for example, in a four-times oversampling filter, three zero values are inserted for every data value output from the disc. This increases the data rate from 44.1 kHz–176.4 kHz. Interpolation is used to generate the values of intermediate sample points, for example, three intermediate samples for each original sample. These samples are computed using coefficients derived from a low-pass filter response.

The spectrum of the oversampled output waveform contains image spectra placed at multiples of the oversampling rate; for example, in a four-times oversampled signal, the first image is centered at 176.4 kHz. Because the audio baseband and sidebands are separated, a low order analog filter can be used to remove the images, without causing phase shift or other artifacts common to high order analog brickwall filters.

Traditionally, D/A conversion is performed with a multibit PCM converter. In theory, a 16 bit converter could perfectly process the 16 bit signal from the disc. However because of inaccuracies in converters, 18 bit D/A converters are often used because they can more accurately represent the signal. Alternatively, low bit (sometimes called 1 bit) D/A converters can be used. They minimize many problems inherent in multibit converters such as low level nonlinearity and zero-cross distortion. Low bit systems employ very high oversampling rates, noise shaping, and low bit conversion.

Also present in the audio output stage of every CD player is an audio deemphasis circuit. Some CDs are encoded with audio preemphasis characteristic with

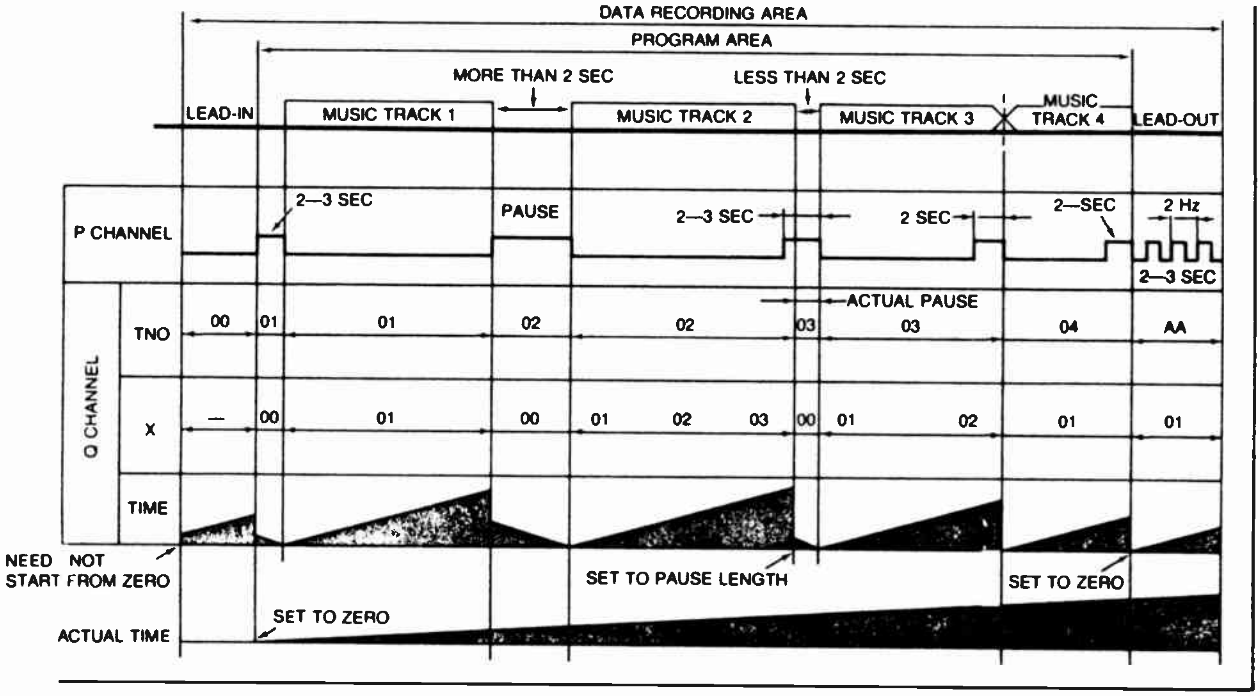


Figure 3.4-12. Typical subcode contents of the P and Q channels. (From Pohlmann, *The Compact Disc.*)

number of channel bits per frame after encoding is 588, comprised of: 24 synchronization bits, 336 ($12 \times 2 \times 14$) data bits, 112 ($4 \times 2 \times 14$) error correction bits, 14 subcode bits, and 102 (34×3) merging bits.

Data Readout

CD pickups use an aluminum gallium arsenide (AlGaAs) semiconductor laser generating laser light with a 780 nm wavelength. The beam passes through the substrate, is focused on the metalized pit surface and is reflected back. Because the disc data surface is physically separated from the reading side of the substrate, dust and surface damage on the substrate do not lie in the focal plane of the reading laser beam and hence their effect is minimized. The polycarbonate substrate has a refractive index of 1.55; because of the bending of the beam from the change in refractive index, thickness of the substrate, and the numerical aperture (0.45) of the laser pickup's lens, the size of the laser spot is reduced from approximately 0.8 mm on the disc surface to approximately 1.7 μm at the pit surface. The laser spot on the data surface is an Airy function with a bright central spot and successively

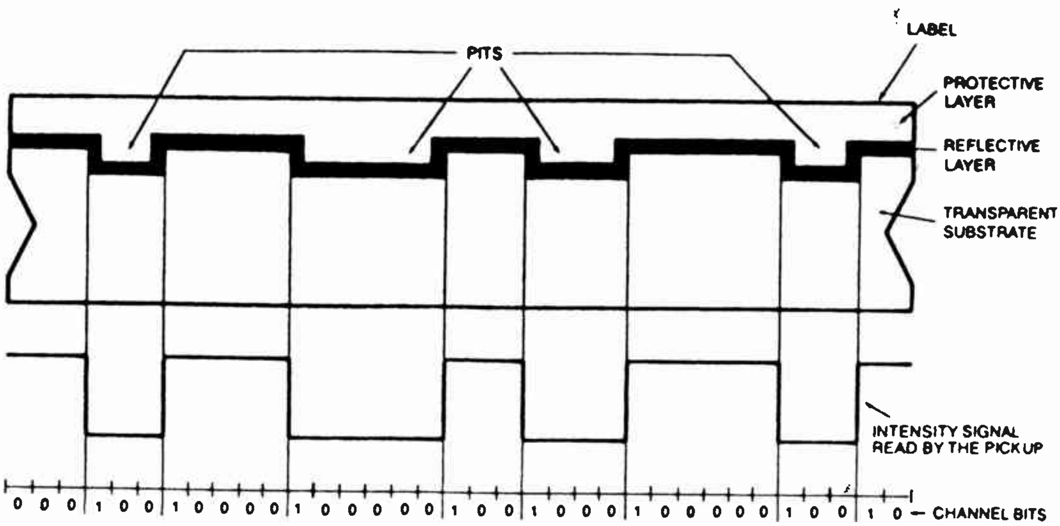


Figure 3.4-13. Channel bits as represented by the pit structure. (From Pohlmann, *Principles of Digital Audio.*)

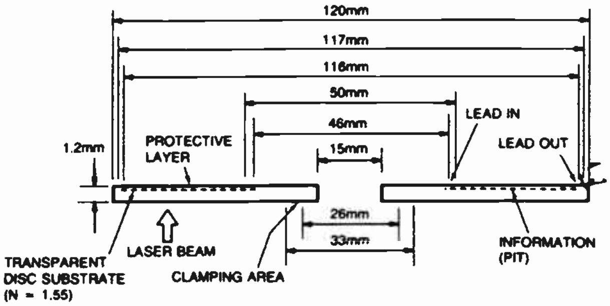


Figure 3.4-11. Compact disc physical specifications. (From Pohlmann, *The Compact Disc.*)

The disc rotates with a constant linear velocity (CLV) in which a uniform relative velocity is maintained between the disc and the pickup. To accomplish this, the rotation speed of the disc varies depending on the radial position of the pickup. The disc rotates at a speed of about 8 rev/s when the pickup is reading the inner circumference, and as the pickup moves outward, the rotational speed gradually decreases to about 3.5 rev/s. The player reads frame synchronization words from the data and adjusts the speed to maintain a constant data rate.

The CD standard permits a maximum of 74 minutes, 33 seconds of audio playing time on a disc. However by modifying encoding specifications such as track pitch and linear velocity, it is possible to manufacture discs with over 80 minutes of music. Although the linear velocity of the pit track on a given disc is constant, it can vary from 1.2–1.4 m/s, depending on disc playing time. All audio compact discs and players must be manufactured according to the *Red Book*, the CD standards document authored by Philips and Sony.

Compact Disc Encoding

CD encoding is the process of placing audio and other data in a frame format suitable for storage on the disc. The information contained in a CD frame prior to modulation consists of a 27 bit sync word, 8 bit subcode, 192 data bits and 64 parity bits. The input audio bit rate is 1.41×10^6 bps. Following encoding, the channel bit rate is 4.3218×10^6 bps. Premastered digital audio data are typically stored on a 3/4 in. U-matic video transport via a digital audio processor with a 44.1 kHz sampling rate and 16 bit linear quantization.

A frame is encoded with six 32 bit PCM audio sampling periods, alternating left and right channel 16 bit samples. Each 32 bit sampling period is divided to yield four 8 bit audio symbols. The CD system employs two error correction techniques: interleaving to distribute errors and parity to correct them. The standardized error correction algorithm used is the *Cross Interleave Reed-Solomon Code* (CIRC), developed specifically for the compact disc system. It uses two correction codes and three interleaving stages. With error correction, over 200 errors per second can be completely corrected.

Subcode Data

Following CIRC encoding, an 8 bit subcode symbol is added to each frame. The 8 subcode bits (designated as P,Q,R,S,T,U,V, and W) are used as 8 independent channels. Only the P or Q bits are required in the audio format; the other 6 bits are available for video or other information as defined by the CD+G/M (Graphics/MIDI) format. The CD player collects subcode symbols from 98 consecutive frames to form a subcode block with eight 98 bit words; blocks are output at a 75 Hz rate. A subcode block contains its own synchronization word, instruction and data, commands and parity. An example of P and Q data are shown in Figure 3.4-12.

The P channel contains a flag bit that can be used to identify disc data areas. Most players use information in the more comprehensive Q channel. The Q channel contains four types of information: control, address, data and *cyclic redundancy check code* (CRCC) for subcode error detection. The control bits specify several playback conditions: The number of audio channels (two/four); preemphasis (on/off); and digital copy prohibited (yes/no). The address information consists of 4 bits designating three modes for the Q data bits. Mode 1 data are contained in the table of contents (TOC) which is read during disc initialization. The TOC stores data indicating the number of music selections as a track number, and the starting points of the tracks in disc running time. In the program and lead-out areas, Mode 1 contains track numbers, indices within a track, track time, and disc time. The optional Mode 2 contains the catalog number of the disc. The optional Mode 3 contains a country code, owner code, year of the recording and serial number.

EFM Encoding and Frame Assembly

The audio, parity and subcode data are modulated using eight-to-fourteen modulation (EFM) in which symbols of 8 data bits are assigned an arbitrary word of 14 channel bits. By choosing 14 bit words with a low number and known rate of transitions, greater data density can be achieved. Each 14 bit word is linked by three merging bits. The 8 bit input symbols require 256 different 14 bit code patterns. To achieve pits of controlled length, only those patterns are used in which more than two but less than ten "0s" appear continuously. Two other patterns are used for subcode synchronization words. The selection of EFM bit patterns defines the physical relationship of the pit dimensions. The channel stream comprises a collection of 9 pits and 9 lands that range from 3–11T in length where T is one period. A 3T pit ranges in length from 0.833–0.972 μm and an 11T pit ranges in length from 3.054–3.560 μm, depending on pit track linear velocity. Each pit edge whether leading or trailing is a "1" and all increments in between, whether inside or outside a pit, are "0's," as shown in Figure 3.4-13.

The start of a frame is marked with a 24 bit synchronization pattern, plus three merging bits. The total

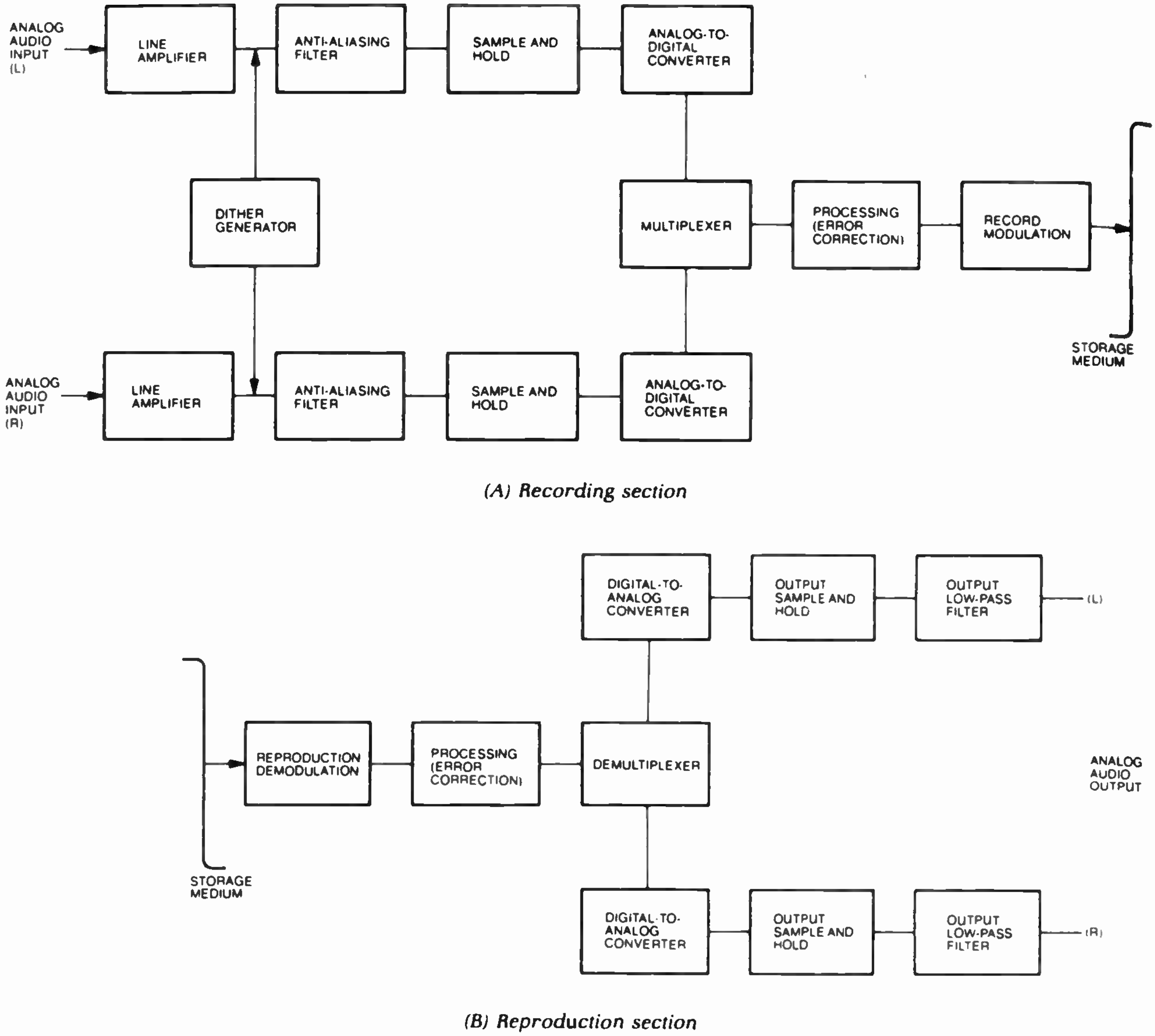


Figure 3.4-10. Block diagram of the recording (a) and reproduction (b) sections of a linear PCM system. (From Pohlmann, *Principles of Digital Audio.*)

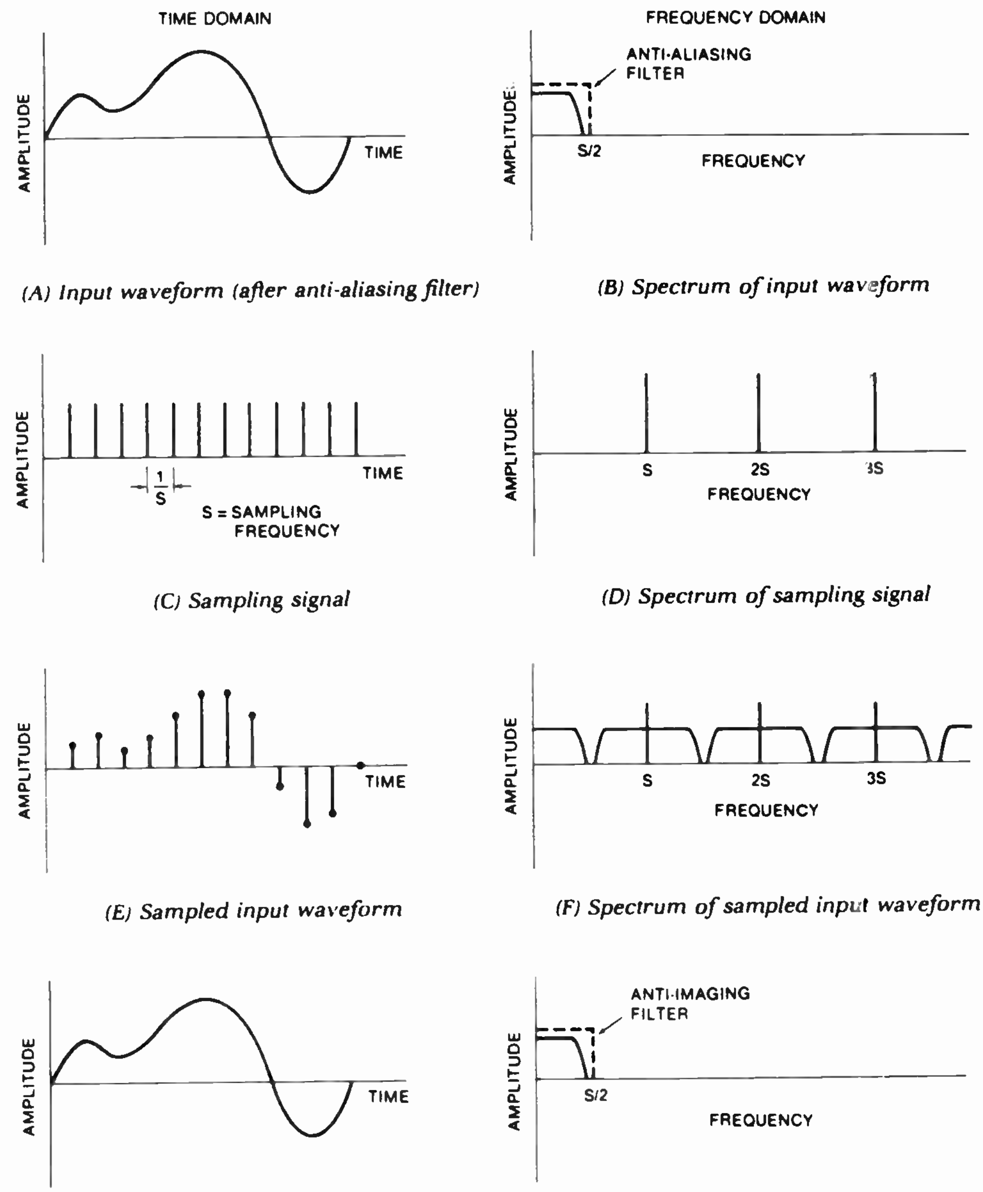
Compact Disk Format

The Compact Disc (CD) format was developed to store up to 74 minutes of stereo digital audio program material of 16 bit PCM data sampled at 44.1 kHz. Total user capacity is over 650 Mb. In addition, for successful storage, error correction, synchronization, modulation and subcoding are required.

Compact Disc Physical Design

The diameter of a compact disc is 120 mm, its center hole diameter is 15 mm and its thickness is 1.2 mm. Data are recorded in an area 35.5 mm wide. It is bounded by a lead-in area, and a lead-out area which contain non-audio subcode data used to control the player's operation. The disc is constructed with a transparent polycarbonate substrate. Data are represented by pits that are impressed on the top of the substrate. The pit surface is covered with a thin metal (typically aluminum) layer 50–100 nm thick, and a plastic layer 10–30 μm thick. A label 5 μm thick is printed on top. Disc physical characteristics are shown in Figure 3.4-11.

Pits are configured in a continuous spiral from the inner circumference to the outer. The pit construction of the disc is diffraction-limited; the dimensions are as small as permitted by the wave nature of light at the wavelength of the readout laser. A pit is about 0.5 um wide. The track pitch is 1.6 um. There are a maximum of 20,188 revolutions across the disc's data area.



(A) Input waveform (after anti-aliasing filter) *(B) Spectrum of input waveform*

(C) Sampling signal *(D) Spectrum of sampling signal*

(E) Sampled input waveform *(F) Spectrum of sampled input waveform*

(G) Output waveform (after anti-imaging filter) *(H) Spectrum of output waveform*

Figure 3.4-9. Summary of discrete-time sampling, shown in the time and frequency domains. (From Pohlmann, *Principles of Digital Audio*.)

When dither is added to a signal with amplitude on the order of a quantization step, the result is duty-cycle modulation that preserves the information of the original signal. The average value of the quantized signal can move continuously between two steps, thus the incremental effect of quantization has been alleviated. Audibly, the result is the original waveform, with added noise. That is more desirable than the clipped quantization waveform. With dither, the resolution of a digitization system is below the least significant bit.

The recording section of a pulse code modulation (PCM) system, shown in Figure 3.4-10(a), consists of input amplifiers, a dither generator, input (anti-aliasing) low-pass filters, sample- and hold-circuits, analog-to-digital converters, a multiplexer, digital processing circuits for error correction and modulation, and a storage medium such as digital tape. The reproduction section, shown in Figure 3.4-10(b), contains processing circuits for demodulation and error correction, a demultiplexer, digital-to-analog converters, output sample-and-hold circuits, output (anti-imaging) low-pass filters and output amplifiers. In most contemporary designs, digital filters are used in both the input and output stages. The output section forms the basis for a compact disc player.

heads the head never touches the tape and thus both head and media life is prolonged. Both crosstalk and signal-to-noise characteristics are excellent in such systems.

Isotropic recording utilizes longitudinal and vertical modes simultaneously. In isotropic recordings the vertical field erases the longitudinal fields near the tape's surface. Thus, the tape is recorded to saturation with longitudinal fields and is multiplexed with vertical fields near the surface. The longitudinal field is structured for dominance at low frequencies and the vertical field carries the higher frequencies. Because the head gaps in isotropic recordings are so minute there is essentially no intersymbol interference because only a small area at the trailing edge of the gap is recorded.

Additionally, because print-through effects are operationally nonexistent much thinner base thicknesses and oxide layers are commonly employed. Coercivity is much higher on digital magnetic media and typically ranges from 800–1500 Oe versus the more typical 300–400 Oe in analog recordings. Thus, digital recordings are deep and robust.

DIGITAL AUDIO RECORDING SYSTEMS[2]

Using the principles of discrete time sampling and quantization, a sampled signal can be processed transmitted or stored and through conversion can reconstruct an accurate representation of the original analog signal.

Discrete Time Sampling

An analog waveform such as an acoustic pressure function in air exists continuously in time over a continuously variable amplitude range. Such an analog function may be discrete time sampled; moreover, the sample points can be used to reconstruct the original analog waveform. This digitization of audio forms the basis for the encoding and decoding of the audio signals in any digital audio format.

The Nyquist theorem states that given correct, bandlimited conditions, sampling can be a lossless process. However the relationship between sampling frequency and audio frequencies must be observed. The Nyquist theorem defines the relationship: if the sampling frequency is at least twice the highest audio frequency, complete waveform reconstruction can be accomplished.

The choice of sampling frequency determines the frequency response of the digitization system; S samples per second are needed to represent a waveform with a bandwidth of S/2 Hz. As the sampled frequencies become higher, given a constant sampling rate, there will be fewer samples per period. At the theoretical limiting case of critical sampling, at an audio frequency of half the sampling frequency, there will be two samples per period. A low-pass filter is placed at the output of every audio digitization system to remove all frequencies above the half-sampling frequency. This is required because sampling, through modulation, generates new frequencies above the audio band. The output filter removes all spectra above the half-sampling frequency. This is summarized in Figure 3.4-9.

By definition audio samples contain all the information needed to provide complete reconstruction. However bandlimiting criteria must be strictly observed; a too high frequency would not be properly encoded, and would create a kind of distortion called *aliasing*. An input frequency higher than the half-sampling frequency would cause the digitization system to alias. If S is the sampling frequency and F is a frequency higher than half the sampling rate, then new frequencies are also created at S ± F, 2S ± F, 3S ± F, etc. An input low-pass filter will prevent aliasing if its cutoff frequency equals the half-sampling frequency. To achieve a maximum audio bandwidth for a given sampling rate, filters with a very sharp cutoff characteristic, *brickwall filters* are employed in either the analog or digital domain.

Amplitude Quantizing

The amplitude of each sample yields a number that represents the analog value at that instant. By definition, an analog waveform has an infinite number of amplitude values, however quantization selects from a finite number of digital values. Thus after sampling, the analog staircase signal is rounded to a numerical value closest to the analog value. The difference between the original values of the signal and values after quantization appears as error.

The number of quantization steps available is determined by the length of the data word in bits—the number of bits in a quantizer determines resolution. Sixteen bits yields $2^{16} = 65{,}536$ increments. Every added bit doubles the number of increments, hence the magnitude of the error is smaller. The accuracy of a quantizing system provides an important performance specification. In the worst case, there will be an error of one half the least significant bit of the quantization word. The ratio of maximum expressible amplitude to error determines the signal-to-error (S/E) ratio of the digitization system. The S/E relationship can be expressed in terms of word length as S/E (dB) = $6.02n + 1.76$ where "n" is the number of bits.

Although a 16 bit system would yield a theoretical S/E ratio of 98 dB, as the signal amplitude decreases, the relative error increases. Consider the example of a signal with amplitude on the order of one quantization step. The signal value crosses back and forth across the threshold, resulting in a square wave signal from the quantizer. *Dither* suppresses such quantization error. Dither is a low amplitude analog noise added to the input analog signal (similarly, digital dither must be employed in the context of digital computation when rounding occurs).

[2] This section contributed by Ken C. Pohlman. *NAB Engineering Handbook*, 8th edition, National Association of Broadcasters, pp. 863–875, 1992.

damage to tape and equipment. Thus, frequent cleaning of heads, guides, capstan, pinch roller required, typically after each recording or playback session is imperative. Freon TF and Ethyl alcohol are considered the most useful cleaners exhibiting very little health hazard, no perceptible solubility on magnetic tapes and virtually no effects on rubber pinch rollers or belts. Unlike Ethyl alcohol Freon TF is not flammable. Careful demagnetization of heads is also required for best performance, typically after each 8 hours of operation. Oils and salts from fingerprints will attract foreign particles and can themselves interfere with reliable head to tape interface.

Hydrolysis is a chemical reaction with water that affects polyester based recording tapes. High temperature and high humidity will accelerate hydrolysis reactions in any polyester based tape stock. However, from roughly 1977–1983 an industry wide polyester binder phenomena, referred to as sticky-shed syndrome, exacerbated the rate of hydrolysis reactivity.

Tapes from the sticky-shed era typically exhibit slip-stick phenomena as carboxylic acid and alcohol are sloughed from the binder as debris products. Tapes of this vintage are frequently unusable due to residue buildup that causes transports to squeal and bind. Fortunately, this phenomenon has been extensively documented and can be reversed temporarily with no apparent damage to the tape recording. The reversal process consists of warming (or *baking*) the tapes in a convection oven at 120°F for 24 hours. The tapes will then be usable upon cooling for several weeks before hydrolysis again sheds sufficient amounts of debris to interfere with transport functioning. Recommended humidity and temperature conditions are shown in Table 3.4-2, below.

Considerations Unique to Digital Recording Systems

Tribology (from the Greek word *tribos* meaning *rubbing*) is critical to high density magnetic recording applications such as digital audio. Thus magnetic tape tribology includes techniques to observe and classify surfaces, friction, lubrication, tape and head wear, contour effects, head-to-disk interface, disk surfaces, sliders, air bearings and contaminants.

Lubricants in magnetic media tend to result in drag forces proportional to media speed and head contact pressure. At higher velocities the coefficient of friction drops to zero at speeds near 80 ips due to the formation of a hydrodynamic air bearing. A friction rise below 1 Hz is due primarily to classic stick slip motions.

Table 3.4-2.
Recommended storage conditions.

Storage Temperature	Maximum Relative Humidity
50° C	39%
40° C	47%
30° C	60%
20° C	79%
10° C	100%

Sliding friction can generate significant heat. The temperature rise at the head/tape interface in helical recorders is typically only 5–7°, but can jump to a flash temperature of 1000° centigrade when an asperity, such as a protruding particle travels past the head.

A great advantage of digital recordings is that system performance is no longer limited by performance of the storage medium. Since transitions are the fundamental language of digital recording systems neither ac bias nor particularly high S/N are required. In fact distorted waveforms are the norm. However, since a massive amount of transition density must be stored for high fidelity audio, higher bandwidth and more precision magnetic emulsions are needed. Linear density, or kilobits per inch is the name of the game. Several techniques are employed to maximum density capabilities, as well as to minimize density requirements.

Error Control and Correction

The need for higher storage densities for digital audio has accelerated research and development in tape composition and magnetic head design. At higher recording densities error vulnerability requires ever smoother recording media and revolutionary designs of recording and playback heads.

Thus, in order to minimize damage and errors due to head to media contact a load-carrying air film is formed at the interface between record head and magnetic media. Physical contact should only occur as the media starts and stops its motion. The air film must be thick enough to conceal any near-contact surface irregularities and thin enough to provide a reliable record and playback signal. Head to medium separation ranges from about 50 nm to 0.3 μm, and the roughness of the head and medium surfaces ranges from 1.5 to 10 nm rms.[1]

Acicular magnetic particles are cigar-shaped particles employed in most magnetic digital recording. Because transitions are the basis of recording, saturation recording is employed and is typically of the traditional longitudinal format. However, for greater storage density the acicular particles can be oriented perpendicularly to the direction of the recording medium's travel. A balance between too low a density which requires excessive tape consumption and too high a density which requires additional error correction to combat drop-outs and intersymbol interference.

Thin film heads are of a substantially different design from analog heads. These heads are manufactured using photolithography to achieve a minute, precise shape. Multi-turn thin-film inductive record heads (IRH) are used for recording but do not have good playback characteristics at slow speeds. However, magneto-resistive (MR) heads are useful due to the output being independent of tape speed. With MR

[1] See Bhushan, Bharat, "Tribology of the Head-Medium Interface," *Magnetic Recording Technology*, Chapter 7, McGraw-Hill 1996.

where they are less noticeable. Clock synchronization techniques are commonly employed in digital audio equipment that correct for mechanically induced effects inherent in the recorder's drive mechanism.

The amounts of wow and flutter that are generally accepted to be just perceptible are:

Speech	.6% peak
Music	.3% peak
Classical Music	.15% peak

Tape Composition

Modern magnetic audio recording tape is of two basic types *particulate* and *thin film*. Particulate media are characterized by magnetic domains coated onto a flexible (polymer film) or rigid substrate (aluminum or glass/ceramic disks). As the name implies, thin film media consist of continuous films of magnetic materials deposited onto a substrate by vacuum techniques. Because thin film techniques yield a smoother and thinner medium than the particulate media, greater recording densities with lower error rates are achieved. Thin film techniques are almost universally used for hard disk storage and are beginning to be used for higher density tape stocks as well.

Particulate flexible media have a magnetic coating on both sides and the substrate is about 76 µm thick. The base film is typically polyethylene terephthalate (PET) film although polyethylene napthalate (PEN) polymers are becoming common for better dimensional stability in thinner tapes. (See Figure 3.4-7.)

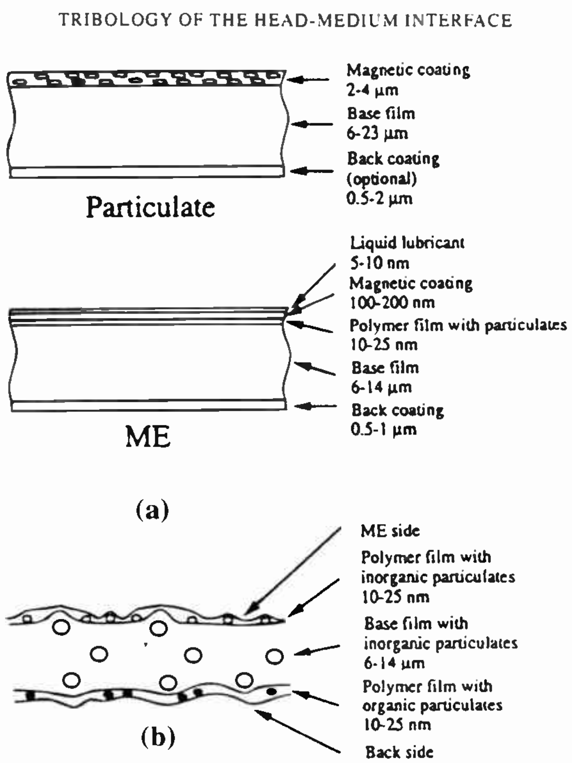


Figure 3.4-7. Sectional views of (a) a particulate and a metal evaporated magnetic tape and (b) coated PET substrate for metal evaporated (ME) tapes. ME is also referred to as thin film flexible media. (From Mee & Daniel, *Magnetic Recording Technology*, 2nd edition, McGraw-Hill, 1995. Reproduced with the permission of the McGraw-Hill Companies.)

Maintenance, Care & Storage of Audio Recordings

Although many analog recordings have held up in good condition for decades they are quite sensitive to permanent physical damage from improper handling, machine malfunction and environmental hazards. Winding tapes *tails out* immediately after complete playback is the most important safeguard in preventing edge damage to audiotapes. Cleanliness and controlled temperature and humidity are the most important factors in preventing environmental damage.

Tape wind or pack must be even to prevent protrusion scatter between layers that will crease and permanently damage tape edges during subsequent playback. Scatter wound tape is susceptible to edge damage from the pressure exerted on flanges during careless handling. For this reason, reels should be handled by their hubs rather than by the flanges. Similarly cinching of layers with actual foldover is possible during rapid acceleration/deceleration from jerky transport operation.

Many professional recorders have a *library wind* mode that operates at a higher than normal operating speed but with constant tension to assure a smooth pack. Tape libraries invariably have professional tape winding equipment that is optimized for gentle handling during higher speed precision winding. At professional libraries preventive maintenance includes periodic rewinding to minimize print-through, depletion of lubrication and to interrupt stiction buildup from adhesive action. Recommended periods between rewindings varies greatly with storage conditions. Tapes stored at 20°C should be rewound every 3,000 hours. Tapes stored at 30°C should be rewound roughly every 300 hours.

Minute particles can cause serious system degradation (see Figure 3.4-8). Static buildup, scraping, scratching of the tape surface and separation in pack and head contact can cause dropouts and permanent

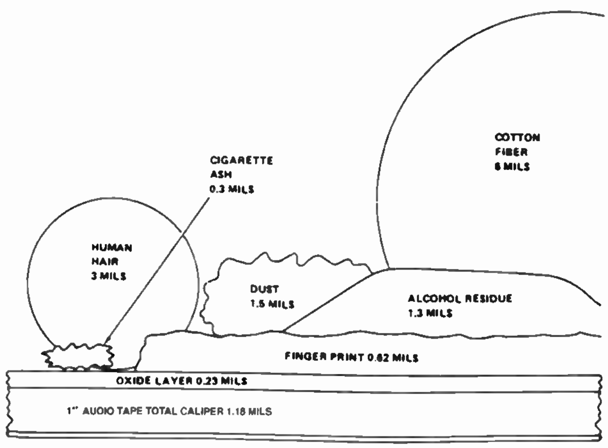


Figure 3.4-8. Relative size of various particulate debris. (From Ritter, *NAB Engineering Handbook*, 7th edition.)

tape move the frequency of maximum print-through lower. Because perpendicular susceptibility in modern audio tape formulations is approximately one-quarter of the longitudinal susceptibility one means to combat the more annoying pre-print effect is to wind tapes oxide out, as is common in European studios as well as standard in cassette tapes.

Velour Effect

As is the case for varying pre- and post-print effects, particle orientation has also been shown to exhibit different high frequency responses in forward versus reverse motion. When particles exist at a nearly right angle to the perpendicular recording field they may not be influenced strongly enough by the recording field. In general the very outermost layer of the coating surface exhibits less remanence. This is referred to as the *dead layer*. The dead layer is actually exploited in some digital recording systems with a longitudinal recording being multiplexed with a vertically recorded signal for greater information storage density. Stronger recordings result when the particle field is closer to agreement with the recorded field as shown in Figure 3.4-6.

Maximum Operating Level (MOL)

MOL_{333} is the maximum level output when the record level has been adjusted to produce 3% harmonic distortion. Some manufacturers use a 5% figure. MOL_{10} is the output between the normal reference level and the tape level at saturation recorded at 10 kHz. Operating levels are expressed in nano-Webers per meter (nWb/m) and are typically between 200–1000 nWb/m for audio tape. In broadcast environments a typical 15 ips setting may employ 320 nWb/m, with 7.5 ips levels set for 250 nWb/m.

Recorded levels depend on the amount of bias current. Each frequency will have an optimum bias current. High bias currents produce a clean, in-depth recording with low harmonic distortion, while a small bias current results in higher high frequency response and high distortion. A universal relationship regardless of tape formulation is that for tape stock with bias adjusted for maximum output at a 5 mil wavelength. One percent distortion is about 10 dB below tape saturation and the 5% distortion level is roughly 5 dB below saturation.

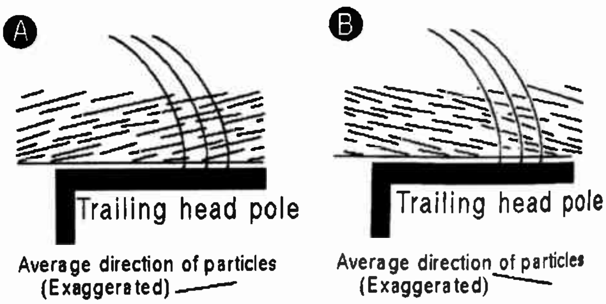


Figure 3.4-6. When particles are perpendicular to the recording field, a weak surface recording results (a); a stronger surface recording results when particles are not aligned perpendicularly to the field lines (b). (From Jorgensen, *The Complete Handbook of Magnetic Recording,* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

Table 3.4-1
Tape speed, coating thicknesses and time constants for audiotapes.

Application	Speed IPS	Coating δ μm	f_δ Hz	$\delta = 1/2\pi f_\delta$ μS	IEC/DIN μS	NAB μS
Studio	30	10	11,395	13	35	10
Professional	15	5	11,395	13	35	50
Home A	7½	5	5,970	27	50–70	50
Home B	3¾	5	2,985	54	90	90
Cassette	1⅞	2.5	2,985	54	120	90

(From Jorgensen, *The Complete Handbook of Magnetic Recording,* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

Coating thickness losses account for a 6 dB/octave increase in the voltage versus frequency response curve at low frequencies. Reproduce amplifiers must employ equalization circuits with a crossover frequency appropriate to the speed and coating thickness (see Table 3.4-1). Because relatively little high frequency energy is present in most music and speech, record current is boosted to achieve a uniform record level at high frequencies. The amount of boost is standardized in the United States according to NAB published criteria and DIN, CCIR and IEC in Europe.

Mechanical Considerations

High quality spooled recording media require precise speed stability and constant tension. In professional machines three motors are typically employed, one for the capstan speed, one for spooling the supply and one for the takeup reel. Most modern professional reel-to-reel analog tape recorders (ATRs) employ constant tension technology to minimize the possibility of stretching the magnetic tape stock. Constant tension is similarly important in achieving good tape pack to minimize cinching, debris pickup, and print-through. Most professional facilities store their tapes *tails out* not only to insure best pack but also to require a rewind prior to subsequent playback to shed any deposits or adhesion effects that can impair playback integrity.

The speed of the capstan must be precise and stable. The capstan's dimension must be perfectly concentric. Since perfection is hard to achieve in practice, mechanical variations that result in speed changes of up to 10 Hz are commonly referred to as *wow*, with speed variations above 10 Hz referred to as *flutter*. Wow is most typically caused by capstan shaft irregularities such as a shape eccentricity or debris pickup such as adhesives or dirt. Motor cogging and layer-to-layer adhesion in a tape pack can also contribute to wow.

Flutter is more typically generated by anomalies in the tape itself. Since magnetic tape has elastic properties, movement over tape guides and heads can cause minute longitudinal oscillations. Resonances set up in the tape medium by tension settings can also cause mechanical vibrations that will be exhibited as flutter. A variety of techniques have been devised to compensate for these effects, including simple masking techniques such as adding idler rollers along the tape path to shift the flutter resonances to higher frequencies

Considerations Unique to Analog Recording Systems

Decades of incremental refinements have been made in magnetic recording techniques. Bias, frequency response, signal-to-noise, print-through, oxide composition, backing media, lubrication and adhesive composition interact with one another and require trade-offs depending on the intended purpose of the recording. A wide range of tape formulations are available, with features specific to the application. Additionally, a wide range of speed, equalization, track widths, noise reduction and level standards have evolved on analog recorders for specific purposes.

Noise, Frequency Response and Bias

Because areas of magnetization increase as the square of the current value, the magnetization and recording process is nonlinear (see Figure 3.4-4). This nonlinearity presents innumerable challenges in reproducing high fidelity audio recordings. Early on it was discovered that when high frequency bias signals are added to the record head, the linearity of the magnetic recording process is improved. This is because the resulting heterodyned signal now moves into the more linear regions of the magnetization curve.

The mechanism of ac bias is familiar to anyone who has manually experimented with producing deep erasures on analog tape stock. The more smoothly the ac field is removed, the deeper the erasure. In practice this principle is applied as the ac bias field decays while the tape stock is moved past the trailing edge of the record gap at a constant velocity. Thus, the high frequency bias field causes the magnetization intensity to ramp down along a predictable hysteresis plot, leaving the tape in a final state that is proportional to the applied audio signal which is heterodyned with the bias signal (see Figure 3.4-5).

The remanence imprint is cleaner with higher frequency bias signals due to limited intermodulation. Thus, the bias frequency is generally set to at least five times the highest audio frequency to be recorded.

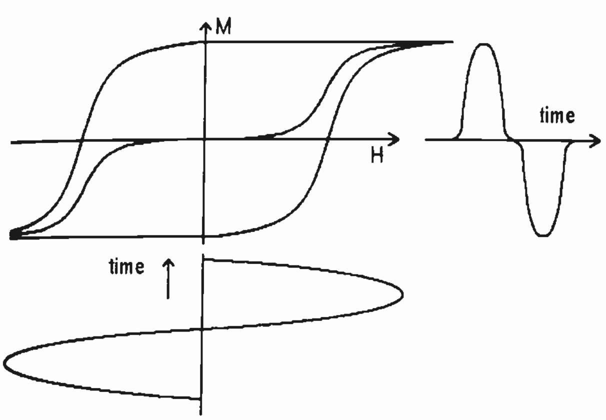


Figure 3.4-4. Zero crossing distortion is shown in a remanent magnetization coating. (From Jorgensen, *The Complete Handbook of Magnetic Recording,* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

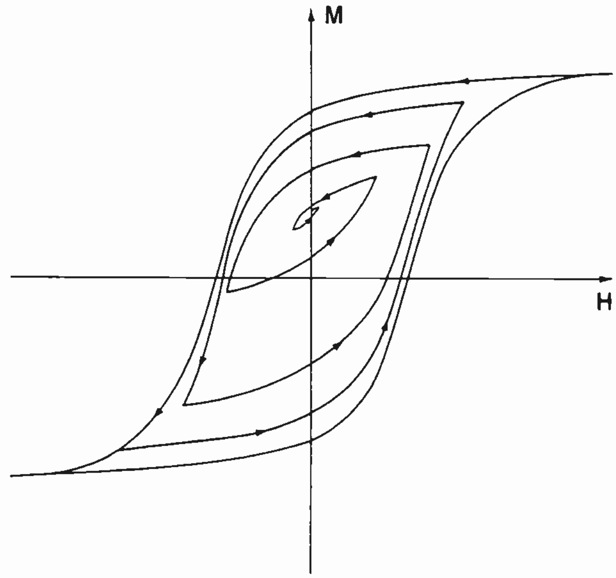


Figure 3.4-5. The principle of AC bias on a moving tape area, showing the decreasing magnetization swing of the high frequency bias signal, leaving the recorded intensity proportional to the superimposed signal field. (From Sharrock, NAB Engineering Handbook, 8th edition, 1992.)

The bias oscillator is also fed to the erase head, and for multitrack recorders a single oscillator is buffered and fed to each gap to prevent intermodulation. Bias current is typically 8–10 times the signal current. Normal bias is defined as the bias field which saturates the tape to its exact coating thickness. Overbias results in short wavelength losses but is frequently employed in professional audio recorders where equalization and higher tape speed maintain high frequency response; thus, even lower distortion is achieved.

Print-through

Print-through affects most magnetic tapes and is of major concern to archivists. As one would expect, print-through is more of an issue in thin base tapes, but coating techniques also affect print-through characteristics. Fe_2O_3 particle sizes can be problematic if too small or if oriented to be susceptible to magnetization levels in the adjacent contact layer. In some early oxide formulations 4% of the particles were small enough to be susceptible to magnetization by adjacent layers. Longer storage times as well as higher temperatures caused print-through to be increased. Typically, both a pre-print and post-print can be observed on each side of a strong signal in the quiet pauses around that signal.

Pre- and post-prints are typically of different magnitudes because the magnetization effect will be additive for one layer and subtractive for the other. When the wavelength of the recorded signal is equal to total tape thickness print-through is greatest. In practice a frequency of 1200 Hz on 2 mil tape at 15 ips combines for greatest print-through. Because this falls at the point of greatest aural sensitivity, methods to combat the effect are important. Slower speeds and thinner

Figure 3.4-2. Early German Magnetophone brought to the USA after capture by officers in the U.S. Army Signal Corps. (From Jorgensen. *The Complete Handbook of Magnetic Recording,* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

modern media to embrace digital techniques, professional archivists continue to prefer and insist on analog recordings due to known fault tolerances and stable format standards that will undoubtedly be readily available for the foreseeable future, even if they are fast losing favor in daily production.

MAGNETIC AUDIO RECORDING OVERVIEW

Domain theory is the basis for explaining ferromagnetic phenomena. In 1907, Pierre Weiss first posited the existence of *atomic magnets*, although Ampere had earlier theorized that magnetism in lodestones might be due to some molecular circulating currents inside the material.

At room temperature there are three materials where there are unpaired electron spins: four electrons in iron, three in cobalt, and two in nickel—precisely the materials in use in magnetic recording technology. These unpaired electrons will undergo exchange between adjacent atoms resulting in quantum mechanical forces of exchange. When the exchange value is positive, neighboring electron spins align and produce strong magnetization levels inside the ferromagnetic material. The spins will align themselves inside small volumes, called *domains* that can be as small as 1 μm or as large as 2 cm. At temperatures approaching absolute zero (0 degrees Kelvin) spin alignment becomes seemingly uniform and magnetization is highest. At higher temperatures thermal agitation causes excursions in the angle of mutual orientation as the exchange forces lose control. The temperature at which the material loses ferromagnetism and becomes diamagnetic is called the *Curie Temperature*. Upon cooling the material becomes ferromagnetic again, but has no memory of its prior magnetization. Typically, between 10^{12} and 10^{15} atomic moments comprise each magnetic domain.

The direction of magnetic domains varies from domain to domain, with net overall magnetization being zero in perfectly virgin ferromagnetic material, such as deeply erased audiotape. Domain behavior has been observed by applying a colloidal iron powder on the polished surface of a sample and viewing the resultant *bitter patterns*.

Magnetic fields are expressed as H, with flux lines shown by the symbol Ø. The intensity of the magnetic field is expressed in terms of flux density B and varies with the number of field lines. Flux density is measured in weber/m^2.

When a magnetic field is strong enough to realign the magnetic domains, even after the field is removed, the amount of magnetism remaining is called *retentivity* or *remanence*. The field required to oppose and reduce the retentivity to zero, or in other words to erase the magnetic field is called *coercivity* and is expressed as H_c.

Coercivity is measured in Oersteds (Oe), with typical analog tape stock requiring about 370 Oe for full erasure. Digital open reel tapes require over 700 Oe and metal particle tapes between 1,200–1,500 Oe for full erasure.

Hysteresis is the lagging of the magnetizing effect behind the magnetizing force as illustrated in Figure 3.4-3. Flux density (B) increases with the field intensity (H), but in a non-simultaneous fashion. As field intensity increases flux density increases reasonably uniformly until approaching saturation beyond which no further flux density is produced. Reducing the field intensity reduces the flux density, but note that when the field intensity is reduced to zero the flux density remains above zero. This is due to the magnetic material's *retentivity*. Only by applying an opposite polarity field of sufficient intensity will the flux density again reach zero. Hysterisis plots are unique for specific media samples due to vagaries in impurities, crystal structure, slip planes and stresses.

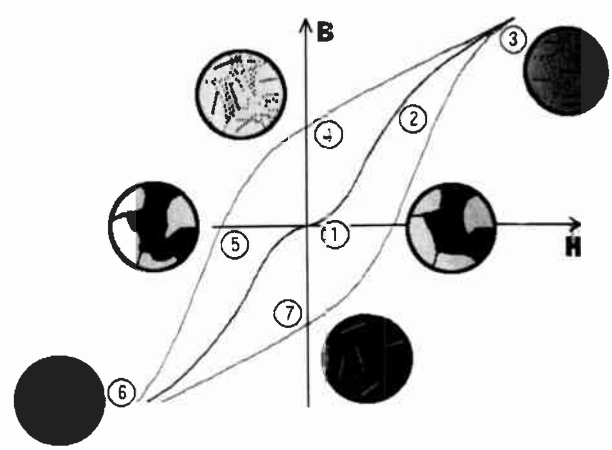


Figure 3.4-3. Depiction of a typical hysteresis loop. 1, 2, 3 represent the original magnetization curve, with 3, 4, 5, 6, 7, 8, 3 representing the hysteresis loop. (From Jorgensen, *The Complete Handbook of Magnetic Recording.* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

3.4
AUDIO RECORDING SYSTEMS

MICHAEL STARLING
NATIONAL PUBLIC RADIO, WASHINGTON, DC

INTRODUCTION

Until just over 100 years ago all sound was *live* sound. It was only with the harnessing and commercial use of electricity that amplification was developed and subsequently led to the widespread ability to store and reproduce sound waves. The faithful reproduction of sound waves is the goal of audio recording systems.

For most of the 20th century, analog magnetic tape has been the broadcaster's principal medium for audio recording. Magnetic recording is an interdisciplinary field of physics, chemical and material sciences, electronics and mechanical engineering. Current practice and research is focused on the implementation and standardization of digital recording techniques using magnetic, optical and even static storage media. The massive storage requirements for digital audio have led to significant advances in perceptual compression techniques and high-density research and development.

History

Audio recording preceded and helped fuel the introduction of broadcasting. The earliest recorded audio was Edison's 1877 cylindrical gramophone employing a constant velocity vertical recording groove. The gramophone's cylindrical media mandated that each recording be a master and stymied mass production. The flat disc recordings that dominated recorded audio for the next half century permitted impression duplication in mass quantities.

Early broadcast use of recorded media exploded in the late 1920's with the introduction of Berliner's mass-produced, laterally recorded flat disc. This development coincided with the rapid proliferation of AM broadcasting. Among the first actions of the Federal Radio Commission in 1928 was the deletion of several stations due to their heavy reliance on airing commercial records which the FRC cited as, "provision of a service which the public can readily enjoy without the service." The new FRC favored original programming and this too stimulated the use of recording lathes at the burgeoning population of radio stations. While many, if not most, early broadcast facilities acquired recording lathes for production of recorded audio, it was not until after World War II that saw the widespread debut of the more forgiving and affordable magnetic tape recording.

Danish telephone engineer Valdemar Poulsen demonstrated a magnetic wire recorder as early as 1898 as shown in Figure 3-4.1. Not until thirty years later did German researchers pioneer magnetically coated, paper-based tape for good quality recorder/reproducers.

By 1936, German scientists had advanced magnetic based recording using cellulose base-tape and achieved remarkably good sound quality. After the war, the AKG Magnetophone (see Figure 3.4-2) was copied and commercially exploited worldwide. Paper tape backing was immediately abandoned. A host of benefits including portability, immediacy of playback, ease of storage, wide dynamic range, low distortion and freedom from ticks and pops propelled magnetic recording to the forefront in broadcasting. Of paramount significance was the influence of a technology that no longer depended on laboratory processing. This was in modern terminology a democratizing development, much as is the current proliferation of digital audio workstations.

For nearly a half-century magnetic tape recording has dominated broadcast production. Even today audio archivists continue to insist that important original recordings come to them on analog reel-to-reel, rather than any of the multitude of recent digital media formats. Despite the inexorable and accelerating trend for

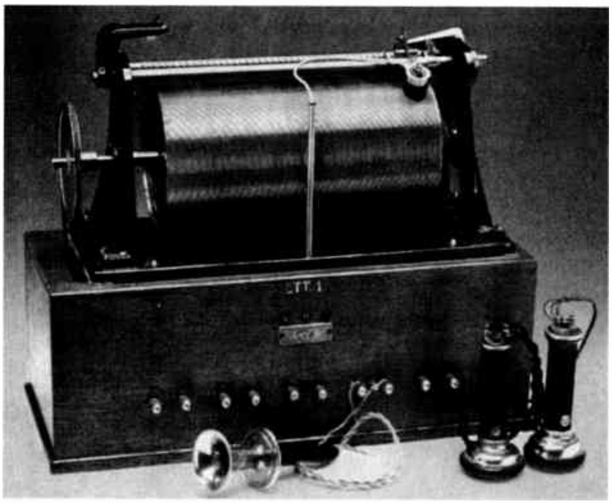


Figure 3.4-1. Valdemar Poulsen's Telegraphone won a Grand Prix Award at the 1900 Paris World's Fair. The device used pole pieces that attached to each side of the wire. Later improvements included a disk surface of 5.25 inches with longitudinal recording on both sides of the disk. (From Jorgensen, *The Complete Handbook of Magnetic Recording,* 4th edition, 1995, McGraw-Hill. Reproduced with the permission of the McGraw-Hill Companies.)

11. Avoid moisture in cables and connectors, particularly where phantom power is being used.

BIBLIOGRAPHY

Burroughs, Lou, *Microphones: Design and Application,* Sagamore Publishing Company, Inc., 1974.

Davis, Don and Carolyn, *Sound System Engineering,* Howard W. Sams & Co., Indianapolis, IN.

Eargle, John, *Sound Recording,* Van Nostrand Reinhold Co., New York, 1976.

Ford, Ty, *Advanced Audio Production Techniques,* Focal Press, 1993

Long, James, "Layman's Guide to Microphone Specifications" *Audio Magazine,* August, 1969.

Long, James, *The Microphone Handbook,* John Eargle, ed., Elar Publishing Company, Inc

"Microphones: An Anthology of Articles on Microphones" *Journal of the Audio Engineering Society,* AES, New York.

Olson, Harry, *Modern Sound Reproduction,* Van Nostrand Reinhold Co., New York.

Sank, Jon R., "Microphones," *Journal of the Audio Engineering Society,* AES, New York, July/August 1985, Vol. 33, No. 7/8.

Tremaine, Howard M., *Audio Cyclopedia,* Howard W. Sams & Co., Indianapolis, IN, 1974.

Woram, John M., *The Recording Studio Handbook,* Elar Publishing Company, Inc., 1982.

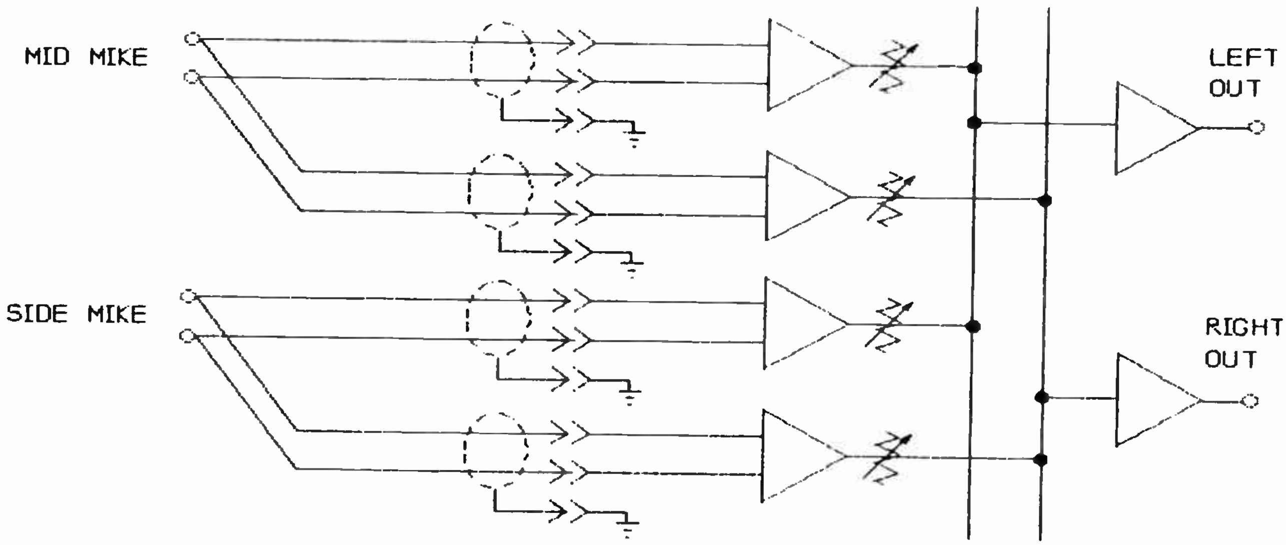


Figure 3.3-34. Mixer used as an M–S matrix.

sound source is placed between the 90° arc of the front capsules of the two bidirectional elements. The stereo sound achieved by this approach can be very natural and mono integrity is well maintained. The Blumlein is more sensitive, however, to ambient noise and reverberation than the M–S and placement is critical.

CARE AND FEEDING

Microphones require a certain amount of care in their handling and storage. Here are some basic factors to consider and tips on microphone care. Misuse, or even some attempts to service or clean the microphone, could affect some manufacturers' warranties. When in doubt, ask or return microphones to the manufacturer's recommended service organization for maintenance.

1. Use a windscreen or pop filter to protect microphone if it is to be subjected to air-borne contaminates such as dust or smoke.
2. A foam windscreen will also protect a microphone from exposure to rain or snow for a surprising period of time. Over time, the cells will fill with water resulting in high-frequency loss and level drop. The foam may be quickly squeezed to reduce the moisture content or a dry screen substituted as required.
3. Foam windscreens will accumulate deposits of dust and other contaminates. The result will be a deterioration of frequency response and, perhaps, even altered polar response. Foam may be cleaned with soap and water. Rinse well to remove all residue. Non-detergent soaps work well.
4. Many microphones may be carefully opened to remove a foam pop filter and sometimes a cloth insert. Do so only in a very clean environment. These filters should be cleaned as detailed previously.
5. Avoid allowing dynamic microphones to be set on work benches or other areas where metal particles or metallic dust may be attracted to their internal magnet structures. Very small particles can work their way onto the diaphragm and alter the response greatly. In some cases, the dynamic microphone can be opened to reveal the diaphragm for examination. Metallic particles may be very carefully removed onto the magnetized tip of a screwdriver. The screwdriver shaft should be steadied on the edge of the microphone case and the tip very carefully lowered to attract particles which would likely be held immediately above the voice coil gap.
6. Avoid subjecting electret condenser microphones to high temperatures. This means do not store in the trunk or glove compartment of a car left in the sun on a hot day. Also avoid leaving the electret microphone on a boom very close to hot lights. The result may be a loss of charge on the capacitor element and a drop in level.
7. Avoid moisture with all microphones but especially with condenser microphones.
8. If given a choice between using mercury or alkaline batteries to power a microphone, remember that mercury cells die much more suddenly than alkalines. The gradual drop in level with an alkaline can be a life saver. Mercury batteries also drop in output level in cold weather. Furthermore, mercury batteries may give off a gas that can corrode the contacts.
9. Avoid unnecessary mechanical shocks. Store in clean, padded enclosures.
10. Moving a condenser microphone from a cold environment to a warm one may cause noise problems from condensation.

Figure 3.3-32. M–S microphone with internal matrix. (Courtesy of Shure Brothers Incorporated.)

A mixer may also be used for deriving L/R information from the M–S pair as shown in Figure 3.3-34.

The M–S technique offers several control capabilities. The first two of these controls may be exercised in either production or post production—another big bonus:

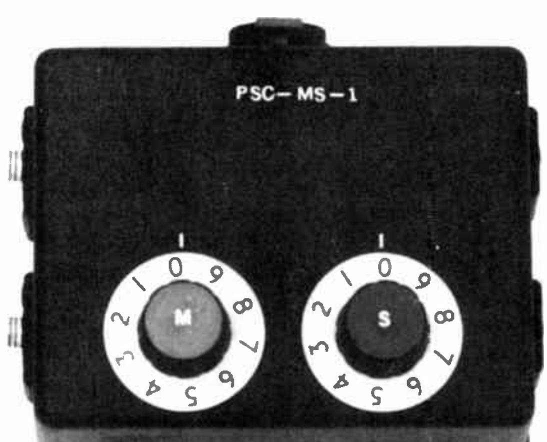


Figure 3.3-33. An M–S matrix decoder. (Courtesy of Audio Services Corporation.)

1. Adjusting the relative levels of the M and the S signals will narrow or broaden the perceived stereo image. This may be done in the field using the M–S microphone's matrix system or you may opt to record the outputs of the M and S capsules on separate tracks, saving the matrixing of them for post production. Matrixing in post will allow the audio perspective to be adjusted to make sense with the video.
2. Panning the M signal off-center may be done to deliberately shift the stereo image. For example, crowd noise at a sporting event may be shifted to appear more closely balanced left and right of a microphone position without moving either the microphone or several thousand fans.
3. Substituting various patterns, from omni to hypercardioid, for the mid microphone will affect the apparent microphone-to-sound source distance as well as the signal-to ambient or reverberant-noise ratio.

Blumlein Miking

The Blumlein technique employs coincident crossed bidirectional elements and, like the M–S, responds to amplitude differences to achieve stereo separation. The

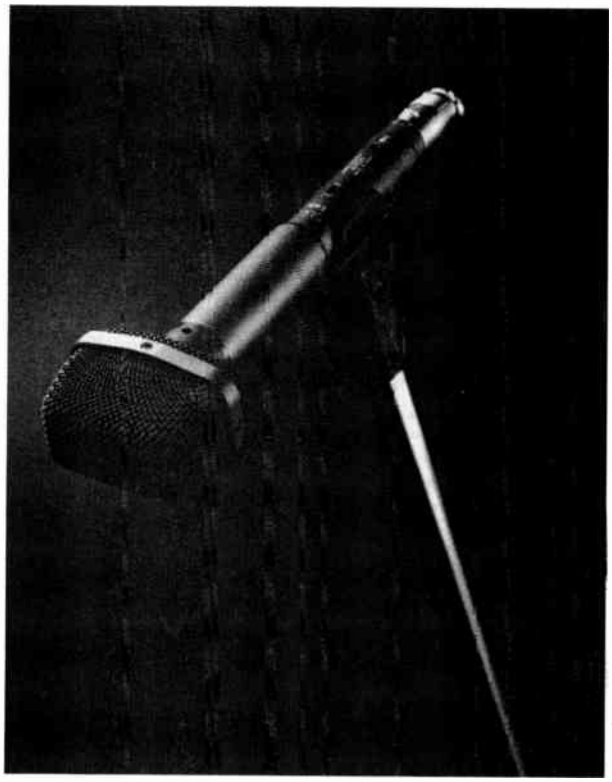


Figure 3.3-29. An X–Y microphone. (Courtesy of Audio-Technica U.S., Inc.)

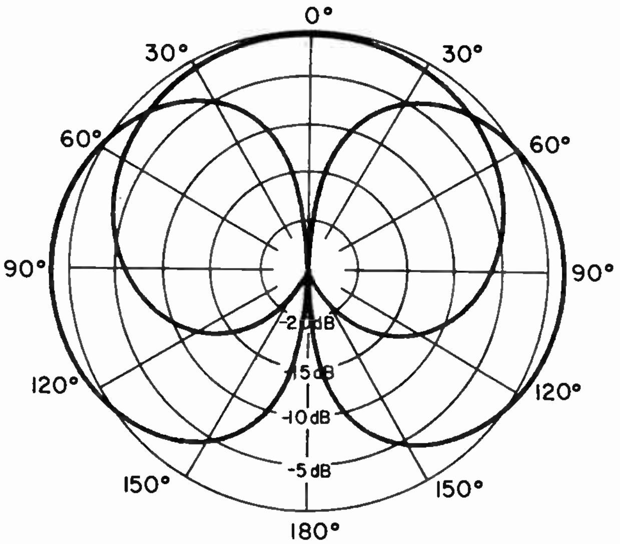


Figure 3.3-30. M–S pattern orientation. (Courtesy of Shure Brothers Incorporated.)

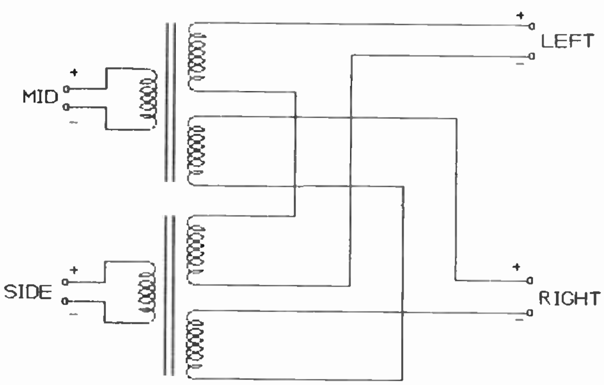


Figure 3.3-31. A passive M-S matrix.

Mid-Side Coincident Microphones

The most versatile of the coincident microphone types for stereo broadcasting is the M–S or "mid-side" microphone, whose polar pattern is shown in Figure 3.3-30. The M–S microphone is a combination of a *mid microphone*, typically a cardioid or hypercardioid, and a bidirectional, *side microphone*. The capsules of the two are placed as close together as possible. A matrixing network combines their outputs and decodes them as left and right channel information, as illustrated in Figure 3.3-31. The information derived from the matrix is nearly identical to that delivered by an X–Y pair, but with some important control advantages.

Sound originating from directly on-axis of the M–S microphone will be picked up by the mid element and delivered equally to left and right channels through the matrix. The mid microphone, which faces the left, with its rear lobe facing right, is insensitive to sound arriving from the center of the stereo stage as the sound is arriving at 90° off axis, where the null is deepest. It is, of course, sensitive to sound arriving from each side.

This is part of the process by which the M–S microphone derives directional cues. Sounds arriving from the left are picked up by the mid and side elements and, because they are in phase, are summed and sent to the left channel. Because the rear of the side element is out of phase with the mid microphone, their sum cannot be used to produce right channel information. Instead, an inverted-polarity version of the side microphone output is mixed with the mid microphone and delivered to the right channel. This processing happens in the sum-and-difference matrix according to the equations:

$$\text{Left} = \text{mid} + \text{side}$$
$$\text{Right} = \text{mid} - \text{side}$$

Commercially available M–S systems offer well-matched capsules, easy operation and considerable control flexibility. The model shown in Figure 3.3-32 incorporates the matrix system and offers a choice of outputs: mid and side or stereo. It also features a three-position control over the degree of stereo "spread" and stereo ambience pickup.

An M–S pair may be constructed using mid and side microphones from the station's collection plus a matrixing system. Several M–S matrices or decoders have been introduced for such use (see Figure 3.3-33).

tial for good stereo without adversely effecting the mono signal with the phase anomalies introduced by spaced microphones. Coincident microphones depend only upon amplitude differences for stereo separation and imaging and provide excellent mono compatibility. There are several coincident microphone schemes including X–Y, M–S, and Blumlein.

X–Y Microphones

The simplest of the coincident techniques is called "X–Y," which crosses two directional microphones so that their patterns meet at their 3 dB-down points (see Figure 3.3-27). The two microphones should be positioned so that one capsule is directly above the other, with the capsules on the same vertical axis. This minimizes any reflection or shadowing of high frequencies that each might contribute to sound arriving on the horizontal plane (see Figure 3.3-28).

An ideal cardioid microphone would have an acceptance angle of 131° and so would be 3 dB down at 65.5° off axis. If a pair of coincident cardioid capsules are rotated apart so as to overlap their polars at 65.5° off each microphone's axis, the resulting angle between the microphones' axes is 131°. If the angle is too great, sound sources at the center of the stereo image are placed farther off axis of each microphone and are thereby attenuated, making them sound as if they are farther away. Similarly, too narrow an angle results in near-center sources sounding louder, or appearing to be closer. Crossing the patterns to overlap at their 3 dB-down points insures that sound arriving from the center of the stereo stage will be summed such as to provide uniform sensitivity from left through center to right.

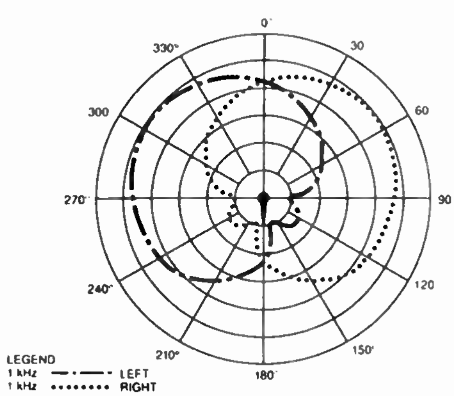


Figure 3.3-27. X–Y pattern orientation. (Courtesy of Audio-Technica U.S., Inc.)

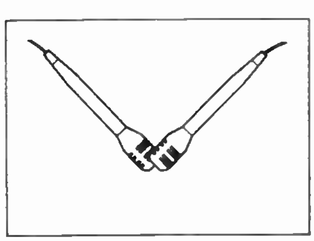


Figure 3.3-28. X–Y capsule orientation. (Courtesy of Audio-Technica U.S., Inc.)

While a 131° angle would be correct for ideal cardioid capsules, the optimum angle for real microphones will likely be somewhat less. It is very beneficial for the microphones used in X–Y to have uniform polar patterns. Because the polars of most cardioid microphones tend to collapse at higher frequencies, it is frequently recommended that X–Y positioning be 90°. This narrow spacing, however, often results in too much overlap of the polars. Stereo separation suffers and center-channel information tends to be brought forward of where it should lie in the stereo image.

The optimum angle for many cardioid elements will be approximately 120°. Experimentation and a thorough knowledge of the polars of the microphones you choose to use will help you get the best X–Y results. Even highly directional shotgun microphones may be used successfully in X–Y, particularly if you are careful to select some of the newer models that have greatly improved polar uniformity. Remember to cross the microphones at the elements, not the ends of the microphones.

A quick check outdoors can be made of the angle adjustment of an X–Y pair using the following procedure:

1. Sum the outputs of the two microphones into a mono audio monitor and provide equal gain for each.
2. Feed pink noise into a small powered speaker about 5 to 6 ft. in front of the shotgun microphones.
3. Rotate the X–Y pair horizontally at their capsules so that you can monitor the pink noise from far left channel to far right.
4. Pay particular attention to the amplitude at center. There should be a smooth transition from left to right channel. If there appears to be a hole in the middle, the angle is too great. If the noise seems suddenly closer at the center, try increasing the angle.

Several X–Y stereo microphones are available that integrate two directional elements into one housing (see Figure 3.3-29). These greatly simplify microphone placement. While easy to use, they should be handled with some measure of intelligence. It may seem obvious that in most cases the axes of the left and right microphones should be near horizontal. Some X–Y microphones, however, hide their capsules in round housings or windscreens that do not permit a quick visual indication of just what is horizontal.

It is not a good practice to use an X–Y microphone for a close-up announcer or reporter application. Even slight side-to-side head movements can cause the voice to shift dramatically from one channel to the other. A good example of a bad application would be to use an X–Y microphone vertically, hand-held, for a stand-up.

microphones are spaced, how closely matched their frequency responses are and the relative levels of the two mixed signals. It should be noted that the terms *phase* and *polarity* do not mean the same thing. Phase refers to a difference in the relative timing of two signals. Polarity refers to the wiring of a microphone or connectors in its circuit and, when reversed, results in a simple shift of 180° in the phase of the signal. Having noted this distinction, in common usage, the terms in phase and out of phase are often used to refer to matters of polarity. Most microphones will be wired to what is sometimes called the "RCA pin count" which is:

Pin #1 shield
Pin #2 high
Pin #3 low

This conforms to IEC standard 268-12 and 268-4. Refer also to EIA-221 (paragraph 3.3) which states that the in-phase terminal shall be the red (or other than black) conductor and that the out-of-phase terminal shall be the black conductor. The terms "in-phase terminal" or "high" (pin #2) indicate the terminal that has a positive voltage present when a positive pressure is applied to the microphone diaphragm.

Checking Polarity

While there are commercially available devices that use a pulse generator to check for polarity reversal, microphones and their cables may also be checked by simply bringing them together and summing their output while speaking into them from a foot or so away. Two microphones which are "in phase" will deliver a higher output under such a test; if they are reversed in polarity, the output should drop noticeably.

Polarity Reversal as a Tool

While inadvertent polarity reversal in a microphone line can result in some very bad audio, deliberate polarity reversal is sometimes employed as a problem solver.

Reducing Background Noise

A pair of microphones may be reversed in polarity to reduce the pick up of ambient noise. This technique is sometimes employed with two microphones in fixed locations, such as in a press box at a sporting event. If these microphones are brought together, a noise-canceling or differential microphone is created. The speaker must now talk into one of the microphones only, virtually in a lip-touching position. Because of the inverse-square law, the amplitude of the voice at that microphone will be much greater than at the other, resulting in reasonable output level. Distant sound will be picked up equally well, however, by both elements and canceled.

STEREO MICROPHONE TECHNIQUES FOR BROADCASTING

As a general rule, in producing stereo audio for broadcast, dialogue should be placed in the center channel, or perceived center, with only music and effects in stereo. Exceptions would be where spatial effects are important to the broadcast material.

Spaced-Pair Microphones

It is of utmost importance to provide high quality stereo audio without compromising mono audio quality. This importance of maintaining compatibility with monaural receivers (or stereo receivers operating in the monaural mode) normally excludes the use of spaced-pair microphone techniques involving omnidirectional or cardioid microphones. Spaced microphones depend upon a combination of amplitude and timing (phase) differences to provide stereo separation. They do not sum well for mono as the very phase differences that aid in separation result in multiple comb filter effects in the mono mix.

Coincident Microphone Techniques

Coincident microphone techniques utilize two microphones whose diaphragms are placed as near to the same point in space as possible. They offer the poten-

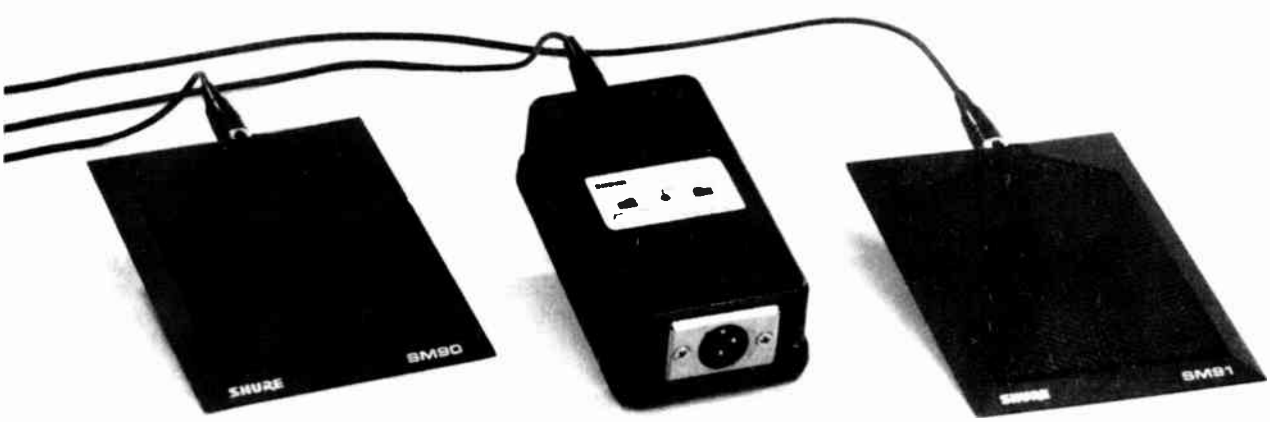


Figure 3.3-26. Cardioid and omni microphones designed for surface mounting. (Courtesy of Shure Brothers Incorporated.)

employed in this manner when redundancy miking is desired for critical applications. When two are used, place one above the other. Normally, only one of these microphones would be open at a time—the second is strictly a backup. Sometimes multiple microphones are used to feed separate systems, such as for house PA, government agencies, and broadcast. Each may still be used as a backup.

Adjacent pairs of cardioid microphones may at times be angled inwardly with their axes crossed and their diaphragms closely spaced. This may be done in order to broaden the acceptance angle of the two microphones while still maintaining some cancellation at the rear. The microphones' close proximity allows their diaphragms to occupy nearly the same point in space, thus reducing sonic time path differences. This ensures that negligible phase cancellation will occur should their output be summed. The same formation is often used as a 2-microphone stereo pickup technique and has the added benefit of good mono compatibility.

Obviously, there are many times when the outputs of two or more open microphones must be mixed. How, then, to avoid phase cancellation? Phase cancellation problems occur when identical signals, at the same or nearly the same amplitude, are allowed to combine at something other than zero phase angle. The result of reducing phase angle error by placing the two pickup elements close together has been shown. It is also possible to influence the amplitude difference between the two combined out-of-phase signals through careful microphone placement. The *3:1 ratio rule* is a good rule of thumb to follow in microphone placement.

To eliminate the problems demonstrated in Figures 3.3-23(a) to 3.3-23(e) by employing the 3:1 ratio rule, D3 must always be at least three times D1. Figure 3.3-25 shows examples of both the violation and enlistment of the 3:1 ratio rule. Subjective tests have shown that an amplitude difference of at least 9 dB between the two signals will reduce phase cancellation to an inaudible level. The 3:1 ratio rule is a means by which this 9 dB minimum difference may be quickly approximated in most multiple-microphone setups.

The amplitude variance desired may also be achieved by judicious use of the mixer's gain or fader controls. Only microphones in actual use should be opened to their normal operating levels; others should be lowered in level or preferably off. Attentive monitoring on an accurate control room speaker system or headphones will reveal audible phase problems, especially with relatively simple sound sources such as speaking voices or most solo instruments.

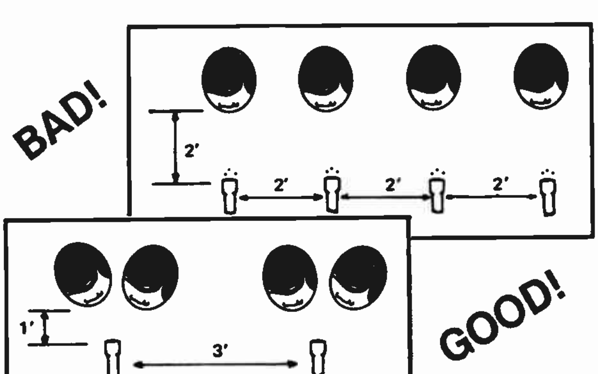


Figure 3.3-25. Obeying and violating the 3:1 ratio rule.

Acoustic Phase Cancellation with a Single Microphone

Acoustic phase cancellation can also occur in a single microphone system when reflected energy from a nearby barrier such as a music stand, podium, table or floor is introduced at the microphone's diaphragm at a sound pressure level within 9 dB of the direct sound. Such problems may be avoided by four different means:

1. Increase the reflected path length.
2. Shorten the direct sound path length.
3. Reduce the reflectivity of the barrier. It may be possible to cover the barrier with an acoustically absorptive material or construct it of an acoustically transparent material. For example, use an acoustically-transparent, visually-opaque screen in chroma keying to eliminate reflections into a weather person's lavaliere.
4. Position the microphone so close to the barrier surface that the direct and reflected sounds arrive at virtually the same time, causing them to be in phase. The latter method also results in higher microphone output because of the additive effect of the two in-phase signals. As discussed in the section on acoustic gain devices, the microphone is operating in (or nearly in) a *half-space environment*. This barrier (boundary layer) miking technique may be employed with omnidirectional or directional elements (see Figure 3.3-26). The omni may be recessed flush into the surface facing out for best performance.

The results of barrier miking will vary because of the following influences:

1. The size of the barrier. The barrier must be large to support low frequency response.
2. The size of the diaphragm. Very small diaphragms may be positioned extremely near the barrier, resulting in less high-frequency cancellation.
3. The reflectivity and resonant characteristics of the barrier.
4. Ambient noise or reverberant energy problems.
5. Reflections from other nearby reflective surfaces.

Microphone Polarity Reversal

Phase cancellation will also occur if the outputs of two microphones, positioned in the same sound field, are combined with their polarities reversed (pins 2 and 3 are reverse wired at one end of one cable). The frequencies canceled, and the degree to which cancellation will occur, will depend upon how far apart the

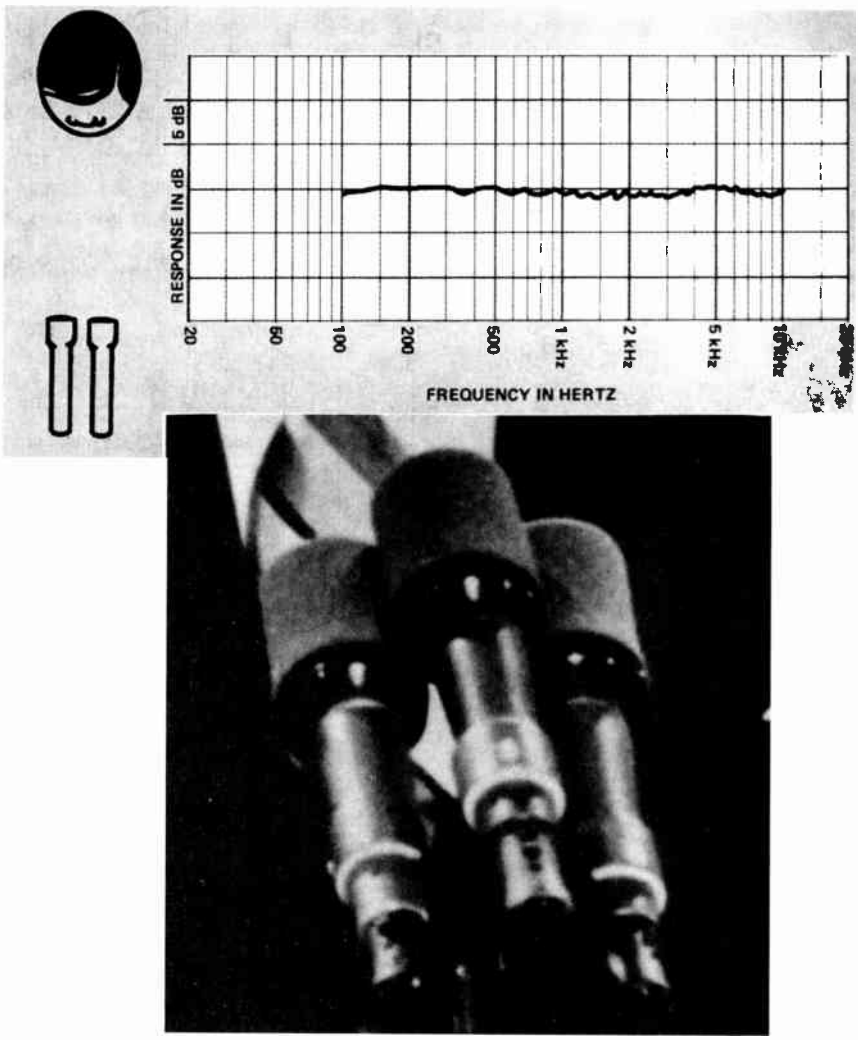


Figure 3.3-24. Redundancy miking. (Courtesy of Shure Brothers Incorporated.)

form (FFT)-derived displays of the actual frequency response of the two microphones combined, with respect to a sound source positioned as shown. The FFT analyzer and its companion microprocessor were used to compare the combined output of two matched, calibrated microphones to the output of one of the two microphones by itself. If no phase cancellation occurred, no trace variations would appear on the X/Y plot, as its plot would be a straight line.

It does not take much study of the response charts to see that, no matter which way the microphones and sound source are oriented, the summed response of the two microphones is poor. These experiments reveal quite graphically what the ear often perceives as a comb-filter or notch-filter effect that sweeps up and down in frequency (and even changes Q) as the variables D1, D2, and D3 change with the movement of the microphones or sound source. In more subjective terms, the resultant sound may be described as hollow, as if the sound is being forced through an empty cardboard tube.

Unfortunately, situations that cause acoustic phase cancellation arise quite frequently. One classic example occurs with a pair of microphones on a podium, spaced apart to provide on-microphone coverage as the speaker turns his head to address all of the audience in front. The curves shown in Figures 3.3-23(b) and 3.3-23(d) are typical of the problems caused by this approach. If the output of these two microphones is summed and fed simultaneously to a house sound system, there probably will be gain-before-feedback problems as well. Add the insistent creeping oscillation of a system on the brink of exceeding unity gain to the already intolerable frequency response of the microphone pair, and you'll have the quality of sound that makes the audience check the credits to see who handled the audio.

The simplest solution to the problems caused by this spaced-pair podium miking technique is to use one microphone only, placing it in front of the person speaking and toward the center of the podium. Figure 3.3-24 shows three microphones immediately adjacent to each other. Two (or three) microphones are often

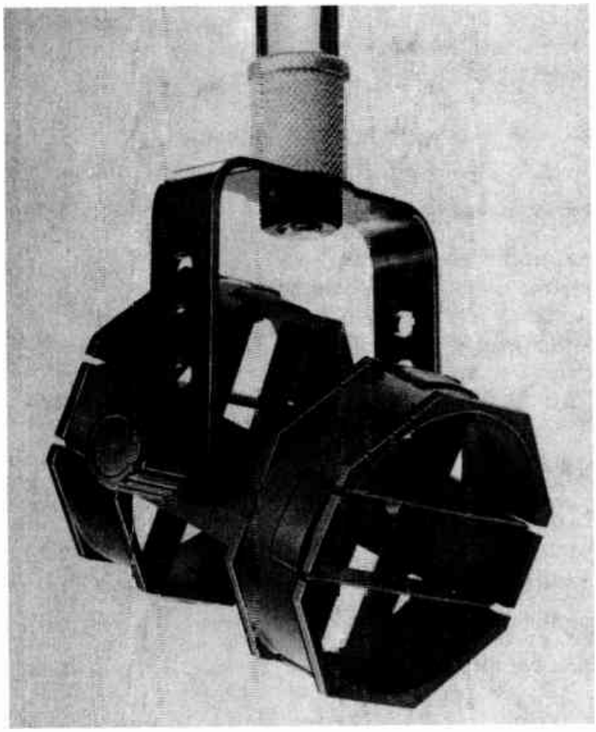


Figure 3.3-22. An external shock mount. (Courtesy of Audio-Technica U.S., Inc.)

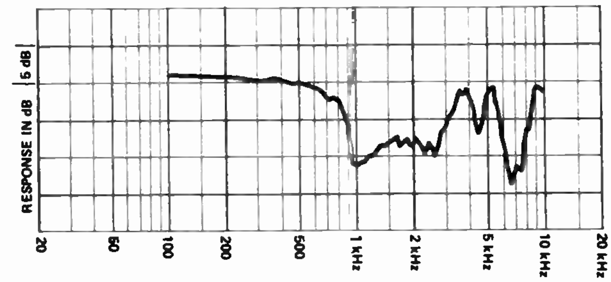


Figure 3.3-23(b). D_1 = 12″, D_2 = 21.6″, D_3 = 18″

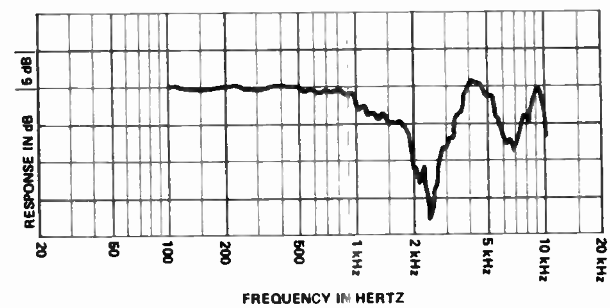


Figure 3.3-23(c). D_1 = 18″, D_2 = 21.6″, D_3 = 12″

to portions of the spectrum that are greatly attenuated. This change in the microphone's apparent frequency response is the result of acoustic phase cancellation. Acoustic phase cancellation may occur when two or more microphones are mixed, or even when a single microphone is subjected to an overdose of reflected sound. Figures 3.3-23(a) to 3.3-23(e) show the severe phase cancellation problems that can result from several typical miking situations.

Although sound arriving at each microphone is identical, originating at the same source, it arrives at the two microphones by paths of varying lengths. This causes a difference in the arrival times and results in phase cancellation of certain frequencies. The curves given for each of these examples are fast Fourier trans-

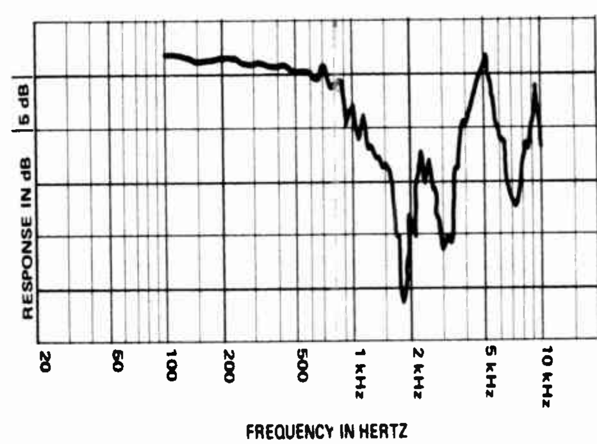


Figure 3.3-23(d). D_1 = 24″, D_2 = 30″, D_3 = 18″

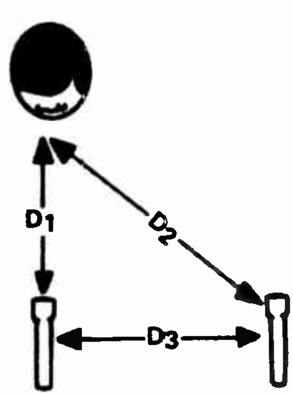


Figure 3.3-23(a). Phase cancellation with multiple microphones.

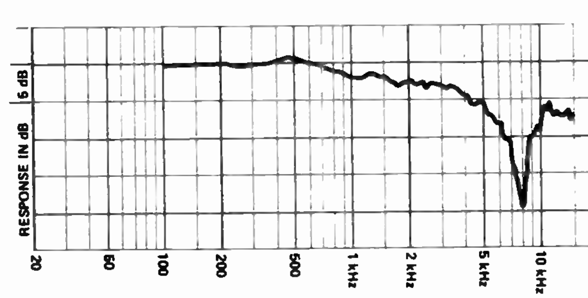


Figure 3.3-23(e). D_1 = 5.6″, D_2 = 6″, D_3 = 2″

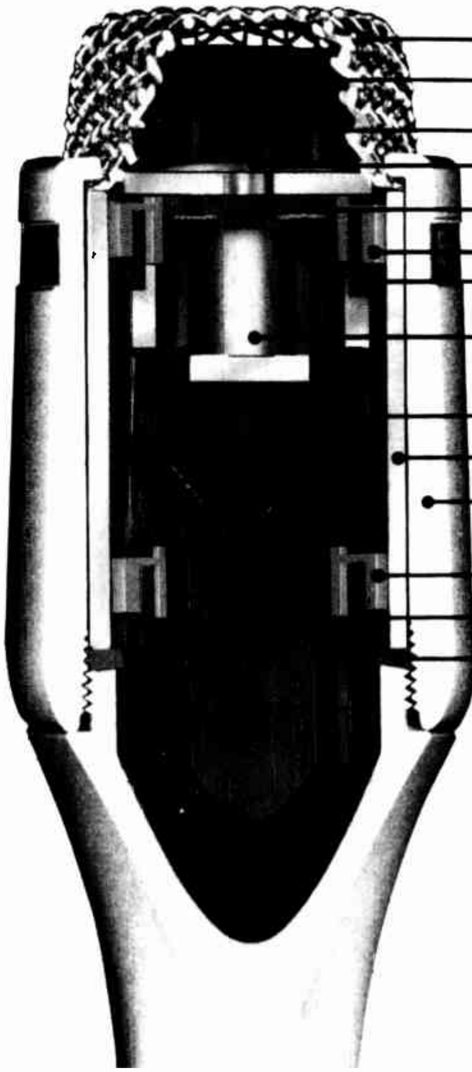


Figure 3.3-21. Cutaway of shock-mounted omni. (Courtesy of Electro-Voice, Inc.)

must cross microphone lines, separate the ac lines so that they cross at different points.

Use twisted pair cable Leads should be twisted inside the microphone and out. The virtually identical positioning this provides for the two conductors within the hum field, and the fact that they are out of phase with each other, will further increase induced hum cancellation.

Use well-shielded cable Installed cable may use a foil shield as flexibility and low flex memory are not a factor. Good stage and field cable, though, normally has a braided shield. Some of the cables which offer the best combinations of flexibility and good shielding use conductive cloth or conductive vinyl under the braided shield.

Follow good grounding practices, avoiding ground loops.

In general, dynamic microphones are much more sensitive to induced hum than are condenser microphones. The voice coil can be a very effective inductor. Hum-buck coils are employed in some designs that lower electromagnetic hum sensitivity by about 20 dB. The hum-buck coil is wrapped around the outside of the motor mechanism and wired in series with the voice coil but out of phase. When both coils are placed into an electro-magnetic field, equal energy is induced onto each. Because they are out of phase with each other, the offending signal is cancelled.

Transformers located within microphones should be avoided if electromagnetic hum is a possible problem. Some transformers are constructed with hum-bucking characteristics, however, to greatly reduce their hum induction potential.

In severe problem situations, operation at line level rather than microphone level may be required.

Acoustic Phase Interference

Another miking problem relates not to sounds that are added to the output, such as popping or hum, but

Figure 3.3-20. Zeppelin-type windscreens for shotgun microphones. (Courtesy of Audio-Technica U.S., Inc.)

duction of a transducer's sensitivity to such noise, or the improvement of its acoustic-to-mechanical-noise sensitivity ratio, starts with the basic element design.

First, omnidirectional microphones are lower in mechanical noise than comparable directional systems. Second, condenser microphones, because of their inherently lower diaphragm mass, are superior in this respect to dynamics. Also, as the size of a dynamic system is reduced (such as in dynamic lavalieres or some small hand-held systems) the acoustic sensitivity tends to decrease while the mechanical-noise sensitivity may remain fairly constant. When the resonant frequency of the microphone's mechanical sensitivity is low, a high-pass filter may be used.

Microphone elements are, of course, often internally shock-mounted by the manufacturer to avoid the transfer of noise from the case to the element. Lowest noise is achieved through the combining of omni or even omni condenser systems with internal shock mounts, as illustrated in Figure 3.3-21.

External shock mounts are often employed in stand or boom-mounted microphone applications. Properly designed shock mounts allow excursion on-axis, or perpendicular to the diaphragm plane. Excellent mechanical isolation may result from using an external shock mount on an internally shock-mounted microphone. Provided that the resonances of the two systems are dissimilar, they will stagger, increasing the effective isolation (see Figure 3.3-22).

Another method of reducing mechanical noise is to raise the resonant frequency of the mechanical drive system. An example would be that of bracing wooden platforms, tables or lecterns to eliminate the very audible, drum-like sound produced when they are struck. The use of very high-density materials for microphone support systems will result in a higher resonant frequency. A microphone stand set onto concrete or into sand gains advantage from this density.

Mechanical noise transfer to the diaphragm may also be reduced through decoupling the diaphragm from tensile forces, converting them to lateral forces. This may be illustrated as follows:

Select a microphone that has some noticeable handling noise problems and plug it into a talk-out system, raising the gain until the handling noise is evident. Now hold the microphone face up (diaphragm horizontal), with the cable hanging straight down, and tap on the cable. This should produce a low thump. Next rotate the microphone 90° so that the cable is hanging at a right angle to the microphone axis. Tap the cable again and the thump should be all but gone. The mechanical drive that takes place longitudinally on the cable is no longer easily transmitted to the diaphragm in the direction of its compliance. Output, therefore is greatly reduced. This principal may be applied to custom hardware designs and the dressing of cables as they enter stand or boom-mounted microphones. A loop of cable or a small coiled cord lowers mechanical noise transfer by this method. Direct interconnection of boom cabling and a shock-mounted microphone can create a direct mechanical short between the two, bypassing much of the shock mount's effectiveness. A coiled cord may be employed between these two points to eliminate this direct path.

AC Hum Rejection

Microphones may also be sensitive to noise induced in them and their cables by electromagnetic or electrostatic radiation. This may be the result of proximity to power transformers, fluorescent ballasts, high-voltage ac lines, SCR dimmers, and so on. The following points should be considered in attempting to avoid or eliminate induced ac hum:

Insure that lines are balanced low impedance The higher the impedance of the microphone, the greater the voltage of the electrostatically induced hum. The balanced line insures that nearly equal hum will be induced on each conductor. Little differential is seen, then, at the amplifier input, resulting in common mode rejection of the hum. Pins 2 and 3 of the 3-pin microphone connector should carry the signal, with pin 2 most often "high." This would mean that positive pressure on the diaphragm produces a positive voltage on pin 2. The shield connects to pin 1 (ground).

Route cables with caution If possible, do not run low-level cables near high-voltage lines, avoiding especially long parallel runs of adjacent microphone and ac power cables. When such cables must cross, they should do so at right angles. If more than one ac line

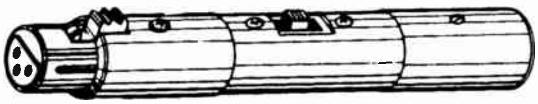


Figure 3.3-19. Microphone in-line attenuator. (Courtesy of Audio-Technica U.S., Inc.)

condenser microphones. A typical in-line attentuator is shown in Figure 3.3-19.

Durability and Reliability

Modern design and manufacturing techniques have provided the broadcaster with a variety of very durable microphone types. There are a few concepts that generally apply to a microphone design's potential durability, including the following:

The moving coil dynamic system offers more inherent durability than condenser or ribbon designs, although both of the latter systems have been employed recently in products that exhibit impressive durability. Also, if dropped on concrete, a lighter condenser microphone hits the surface with less momentum than its heavier dynamic counterpart.

Omnidirectional microphones are more easily built to be durable than are more complex directional designs.

Internally shock-mounted microphones may present a compromise in durability relative to rigidly mounted, nested designs.

Reliability is influenced by a variety of factors including design, construction, and quality control as well as the care that is taken in the use, transport, and storage of the microphone.

Voiding and Avoiding Noise Problems

Other unwanted signals include wind noise, "P-pop," mechanical or handling noise, AC hum and radio frequency interference (RFI). The reduction or elimination of each of these can be handled both through microphone design and user technique.

RFI problems can usually be traced to a point in the low level circuitry at which the signal leads are unbalanced, high-Z, or both. Condenser microphones, for example, may sometimes be sensitive to RFI at or around their impedance converter. In such an event, the manufacturer should be consulted for low-pass modifications or information.

P-pops may be reduced through several approaches:

Use an omni if possible Directional microphones are much more prone to P-pop problems than are omnidirectional ones.

Position the microphone out of the area of the breath blast In an announce application, speak across the microphone: on-axis but at a slight angle to the diaphragm. Stand-ups and hand-held interview miking should be done with the microphone capsule below the axis of the mouth.

Use a pop filter This is often the same as the manufacturer's windscreen. Avoid using foam filters that are not designed for the particular microphone model. If such a requirement arises, test the combination carefully for frequency response and directional characteristics before putting it into service. Windscreen/pop-filter foam is specially designed, reticulated foam that comes in a variety of densities. Even very acceptable open-cell foam may be too thick for use on some microphones. Nonreticulated foam (such as nerf balls) will roll-off high frequency response and alter the polar patterns of directional microphones.

Fashion a pop filter For radio and other off-camera miking, a piece of silk may be suspended a short distance in front of the microphone diaphragm. This is easily accomplished using a frame made of small, wooden hoops—the type used for embroidery. The outer of these concentric hoops expands for removal from the other ring. Surround the inner hoop with nylon hose material, pulling the fabric to one side where it may be sewn together with thread, and the excess cut off. Replace the outer ring and tighten it with the screw mechanism supplied. Attach the pop filter to an arm that will keep it positioned in front of the microphone. Adjust the distance between the two for best performance.

Use a high-pass filter Most of the disturbing explosive energy of a P-pop is very low in frequency. Try using a very abrupt high-pass (low-cut) filter in the microphone line. Rolling this energy off before it gets to the mixing board or recorder input will further reduce distortion in the audible range. Wind noise may be dealt with similarly to P-pop:

1. Omnis are preferred for low wind noise.
2. Use a properly designed windscreen. Most are made of reticulated foam. Superior results with shotgun microphones can be attained by using a well-engineered fabric/mesh cylindrical screen which provides an air space between the material and the microphone (see Figure 3.3-20). To handle severe cases (gale-force winds) special fur-like socks are available to wrap around the tubular windscreen. While this will result in some performance trade-offs, recordings made under such conditions are typically not intended for critical listening. Windscreens must cover all openings to the element: front and rear.
3. Use a high-pass filter.
4. A microphone with a limited low frequency response will help minimize wind noise. Extended-response condenser microphones can produce very high outputs of infrasonic energy when panned or when air around them is moved by air handling systems, etc. The result may be preamp overload or undesirable compressor or limiter action. Again, windscreens and or high-pass filters may solve these problems.

Handling or Mechanical Noise

The problem of mechanical, nonacoustic noise is one that plagues the user whether the microphone is hand held, body worn or hardware mounted. The re-

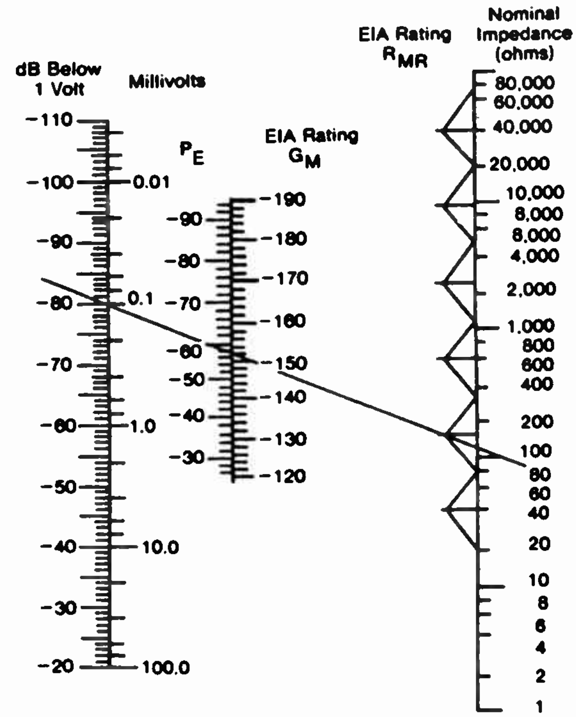


Figure 3.3-17. Nomograph: EIA sensitivity rating.

Figure 3.3-18. Microphone preamplifier. (Courtesy of Shure Brothers Incorporated.)

ance, ranging typically from 50 to 600 Ω. Dynamic moving-coil microphones achieve their low impedance by either a low-Z voice coil winding or a transformer. Condenser microphones use an impedance converter circuit to step-down the capacitor's high-Z output.

Low impedance offers the advantages of low susceptibility to hum and electrical noise pick-up, and the ability to use relatively long lines with a minimal loss of level or high-frequencies. Unlike matching power amplifier impedances to speaker systems, which may be desired for best power transfer, microphones need load impedances on the order of 10 times their internal impedance or even higher. This assures maximum voltage transfer. A microphone that looks like a resistive source of 150 Ω, looking into a load resistor of 150 Ω, for example, will suffer a 6 dB voltage drop compared to an open-circuit connection.

Dynamic Range

The difference between a microphone's own self noise and the maximum sound pressure level (SPL) it can handle is its *dynamic range*. In many field applications, ambient noise provides sufficient masking to make the self-noise specification of minor interest. The importance of this specification increases, of course, as greater working distances are demanded or ambient noise levels are lowered.

The impedance converter of condenser microphones, like any active circuit, will create some noise. The degree to which it does so will vary greatly from one design to another. The impedance converter design also determines the headroom, or maximum SPL that the condenser microphone can handle. A maximum SPL of as much as 141 dB is achieved in several high-quality condenser microphones. Such levels may never be encountered by lavalieres worn at a news desk or seldom by shotgun microphones that are more often attempting to pick up weak audio signals.

Dynamic microphones contribute virtually no self noise. When greatly amplified, only the noise of the thermal agitation of air molecules is detected. While this is very low in level, the dynamic microphone does not automatically rank as the first choice in a low-noise system. Because the output level of the dynamic is often lower than that of a condenser system, the user may end up working into the noise floor of the preamplifier in order to provide sufficient system gain.

Some new dynamic microphones employ powerful rare-earth magnets to increase the efficiency of their motor mechanisms. Their higher output, while still not as high as many condenser microphones, can provide a considerable S/N advantage over earlier designs.

Properly designed dynamic microphones are practically impossible to drive to audible distortion. The distortion heard when a dynamic microphone is subjected to the lips-touching proximity of a very loud rock and roll vocalist is usually the clipping of the electronics following the microphone. Outputs of 1 V or more may actually be delivered in such applications as rock and roll music. Preamp or amplifier clipping may be avoided by padding the microphone output or adjusting the trim (gain adjustment) of the mixer. Be aware that many mixers offer adjustment only after a gain stage or transformer, either of which may distort before any control is possible. In-line attenuators, or pads, are commercially available that allow the selection of 10, 20, or 30 dB of attenuation. These plug directly into the microphone line. Before using any in-line device, it is a good idea to verify that it is compatible with the powering system being used for

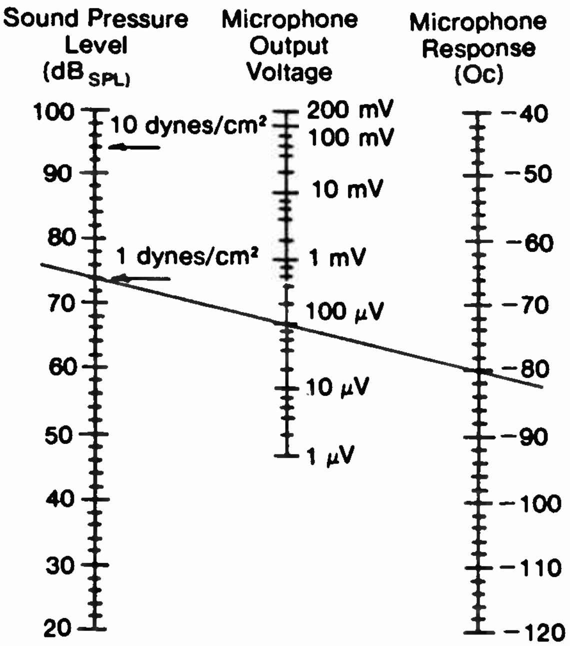


Figure 3.3-15. Nomograph: open circuit voltage rating.

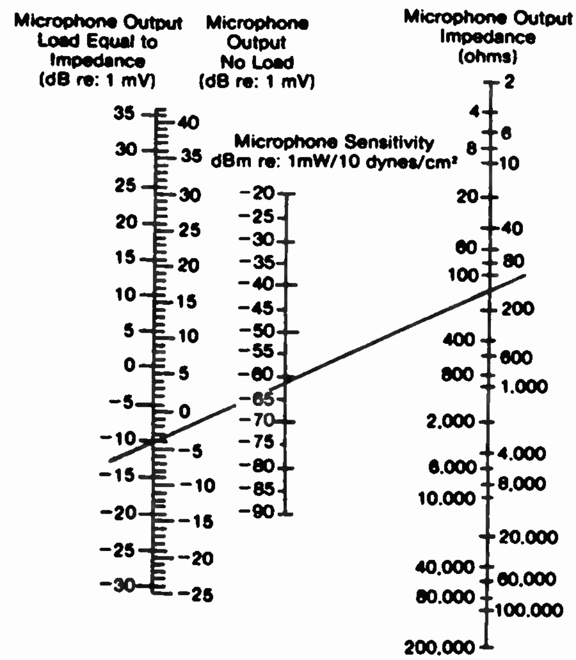


Figure 3.3-16. Nomograph: equivalent power rating.

Power Level

Microphone sensitivity may also be specified in terms of its output power level. This equivalent power level rating takes into consideration the open circuit rating and either the actual impedance of the microphone or the rated impedance. Specifications given would be in dBm (or just dB) referenced to 0 dB = 1 mW/10 dynes/cm² or 0 dB = 1 mW/Pa.

Calculating the power level rating may be done with the following formula:

$$P_E = V_{OC} - 10 \log_{10} Z + 44 \text{ dB}$$

where,

P_E = Equivalent power level
Z = Impedance of the microphone

A nomograph for determining the equivalent power rating is shown in Figure 3.3-16.

EIA Sensitivity Rating

The EIA sensitivity rating is one of those specifications that is sometimes specified but hardly ever used. The formula for determining EIA sensitivity is as follows:

$$ESR = V_{OC} - 10 \log_{10} R_{MR} R - 50\text{dB}$$

where,

ESR = EIA sensitivity rating
V_{OC} = Open circuit voltage in dB
R_{MR} = Center value of the nominal impedance range

Table 3.2-2 may be used for determining R_{MR}. An EIA sensitivity rating nomograph is given in Figure 3.3-17.

Table 3.3-2
Table for determining R_{MR}.

RANGE (in Ohms)	CENTER VALUE (in Ohms)
20–80	38
80–300	150
300–1,250	600
1,250–4,500	2,400
4,500–20,000	9,600
20,000–70,000	40,000

Line-Level Microphones

Line-level microphones designed for remote use incorporate a microphone, preamplifier, limiter, and power supply in one hand-held package. Line level may be required to get the signal above the level of induced noise to overcome resistive losses in long cable runs or merely to level-match to a system that does not provide a microphone preamp at its input. Applications of line-level microphones may include operation into extremely long lines, into unshielded twisted-pair cable, phone lines, or into line-level inputs such as those often found on microwave transmitters.

A line-level microphone is sometimes preferred over a separate microphone and preamp for simplicity of operation. A separate preamp, such as the one shown in Figure 3.3-18, however, allows use of a variety of microphones and offers several powering, control, and connection advantages.

Output Impedance

The impedance (Z) of a microphone is a measurement of its AC resistance looking back into the transducer. Broadcast microphones should be low impedoff-axis sounds originating from the microphone side of the window.

Acoustic Gain Devices

Shotgun microphones increase working distance by ignoring off-axis energy, thereby narrowing the acceptance angle. Some devices actually increase working distance by providing *acoustical gain.* Increased acoustical gain means less electrical gain is required and, therefore, less electronic noise is encountered. [If a satisfactory signal-to-ambient-noise ratio is obtained, the improved signal-to-electrical-noise ratio derived from the use of an acoustical gain device will increase the effective working distance of the overall system.]

The most commonly used acoustic gain device is the *parabola* or *parabolic reflector*, as shown in Figure 3.3-14. The parabolic reflector is shaped so that sound is reflected onto a focal point a short distance in front of the center-point of the dish. An omnidirectional microphone placed at this point receives multiple reflected sound waves in phase, which add to produce significant gain. The response of such systems is very ragged and limited. Low frequency response is extended as the dish diameter is increased. While totally unacceptable for most broadcast applications, the audio quality achieved with the parabolic microphone is often deemed adequate for sound effects pick-up such as at sporting events. In some situations, acceptable quality means the best level of performance possible through practical or affordable means.

A second type of acoustic gain device is the *horn.* Low cost re-entrant horns are often used for talk-back in paging systems. They can be quite directional and are extremely sensitive. Their audio quality as a microphone, however, is no better than a loudspeaker. Horns are sometimes used as microphones in surveillance applications where natural sound is secondary to sensitivity and intelligibility. Installed on the side of a building, the small horn is virtually as inconspicuous as a light fixture and is seldom thought of as a microphone. Some horns are built with 45 Ω voice coils, providing higher output signals to microphone inputs.

Acoustic gain is also realized by using a microphone in the very close vicinity of a large hard, reflective surface. Omnidirectional microphones may be flush-mounted into the barrier, facing out. In this position the microphone is in a half-space environment, or looking into only half the world. The output, for sound arriving on-axis, is increased by 6 dB. As the sound source is rotated off-axis, however, the microphone output drops. At 90° off-axis, the output is down 6 dB, or equal to the omni in free space. The resulting polar pattern resembles a cross-section of a cardioid cut through the microphone at 90°.

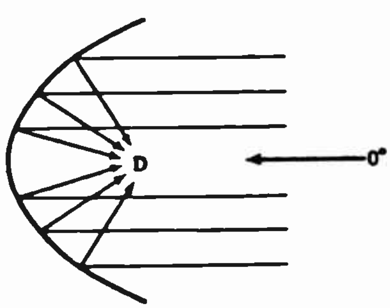


Figure 3.3-14. Parabola principle.

Frequency Response and Distant Miking

Distant miking may result in noticeable or even dramatic changes in the spectrum of the sound being recorded. High frequencies, attenuated by the air, may require boosting to restore a normal sound. Similarly, high-pass (low-cut) filters may prove helpful in reducing low frequency room reverberation or background noise, thereby extending the useful working distance. If possible, any boost in equalization (EQ) should be done in the recording process to minimize noise that would be amplified in post-production equalization. Microphones with rising high frequency response or a rolled-off low end will reduce the need for EQ.

Sensitivity Ratings

One of the greatest sources of confusion on a microphone specification sheet has to be the sensitivity rating. Several different rating systems are currently in use. Table 3-3.1 includes some figures to keep in mind when trying to interpret the ratings.

Open Circuit Output Voltage

Microphones are often specified as having a particular output voltage when looking into an open circuit. In most modern equipment, microphone inputs are at least 10 times the measured impedance of the microphone and may be regarded as an open circuit. Specifications may be given as an actual output voltage or as decibels below one volt at a sound pressure level of 74 dB (1 dyne/cm^2 or 0.1 Pa). These ratings are referred to as *the open circuit output voltage* rating or *open circuit sensitivity.*

The open circuit sensitivity may be expressed in dB by means of the following formula:

$$V_{OC} = 20 \log E_O - SPL + 74$$

where,

V_{OC} = Open circuit voltage in dB (referenced to 1V/0.1 Pa)
E_O = Microphone output in volts
SPL = Actual SPL at the microphone

The nomograph shown in Figure 3.3-15 provides an easy method of calculating the open circuit voltage.

Table 3.3-1
Converting specified parameters

10 dynes/cm^2	=	+94 dB SPL
	=	10 microbar
		1 pascal (Pa)
	=	newton/cm^2

NAB BROADCASTERS

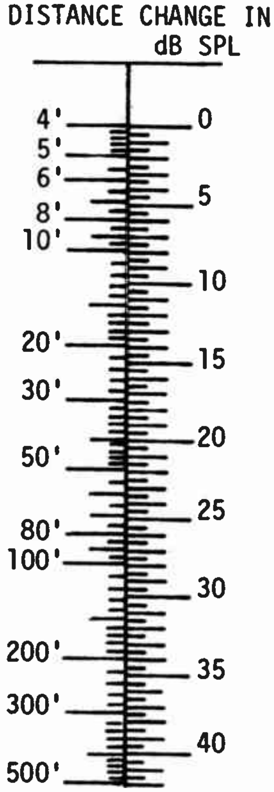


Figure 3.3-12. Inverse square law.

source of one half increases sound intensity at the microphone by a factor of 4 and the SPL by a factor of 2, or 6 dB, as illustrated in Figure 3.3-12. As the input sensitivity control of the mixer or recorder is lowered to compensate for the additional 6 dB available from the now-closer sound source, the microphone, in effect, becomes less sensitive to distant sounds.

Headset Microphones

Headset microphones provide benefits gained from always being a fixed distance from the speaker's mouth. Background noise is reduced (because of the inverse square law) and levels remain consistent. Omni, cardioid, and differential elements are available in headset systems. Cardioids offer the best combination of ambient noise suppression and acceptable broadcast quality.

Going For The Distance

In some distant-miking applications, the effective maximum working distance may be determined by the electronic signal-to-noise ratio of both the microphone and subsequent amplifiers. For example, the selection of an ideal boom microphone for picking up dialogue in a quiet environment, with no reverberation problems, may have little to do with polar patterns. Instead, the end user might look for a microphone with high output and extremely low self-noise. Most often, though, one is concerned with both electronic signal-to-noise and signal-to-ambient noise ratios.

Shotgun Microphones

Effective working distances beyond those afforded by cardioid, supercardioid, or hypercardioid systems may be realized through the use of a shotgun microphone. The shotgun microphone uses a slotted interference tube ahead of the element to provide a high degree of cancellation at the sides. Sound waves arriving on-axis are essentially unaffected by the tube. Sound arriving from slightly off-axis, however, is forced to turn and travel down the tube to the element. This results in numerous out-of-phase conditions being set up in the tube, with cancellation increasing as the microphone is rotated to 90°. The length of the tube will determine the low frequency limit to which this cancellation can be effective.

All things being equal, longer tubes are more likely to maintain their cancellation at lower frequencies than short tubes. Newer shotgun microphone designs from several manufacturers provide superior pattern control using shorter interference tubes than those required by older standards. The new generation of shorter, lighter products are much easier to handle in both fishpole and studio boom applications.

Some shotgun microphones are much less uniform in off-axis response than simpler, more conventional designs (see Figure 3.3-13). Even with their multi-lobed polar patterns, however, the very narrow acceptance angle can often save the day.

Shotguns work best outdoors and in controlled acoustic environments such as well-designed studios. They do not function properly near reflective surfaces. Distant miking down a hallway will not be assisted greatly by the use of a shotgun microphone. A shotgun microphone pointed toward an open window will see all sounds originating on the other side of that window as if they are coming from a source the size of the window. No directional advantage will be realized other than to reduce, perhaps, the level of some nearby

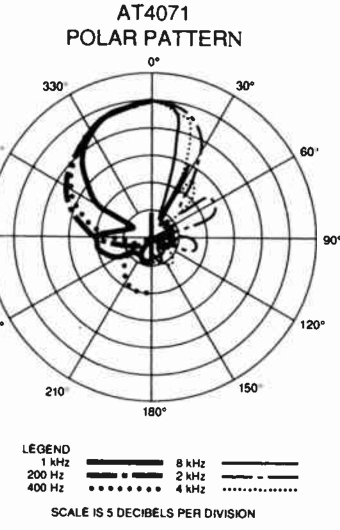


Figure 3.3-13. Shotgun polars. (Courtesy of Audio-Technica U.S., Inc.)

Omni vs. Cardioid

Compared to the omnidirectional microphone, the cardioid has several characteristics that should be noted:

- The cardioid reduces the pick-up of ambient noise and reverberant energy. A look at the comparison chart shows that the *random energy efficiency* (REE) of the cardioid is 0.333 (compared to a REE of 1 for the omni). The random energy efficiency is a measurement that compares a microphone's sensitivity to random (or reverberant) energy to its on-axis sensitivity. While this shows the cardioid to be one third as sensitive to random ambient noise as the omni, remember that discrete sound sources positioned at the null of the cardioid will be attenuated to a much greater extent. This will prove true outdoors where sounds arrive at the microphone directly with minimal reflections. Indoors, the advantage of the cardioid's deep rear null is only appreciated when the microphone is situated within *critical distance* of the offending sound source. Up to the critical distance, the direct sound is greater in intensity than the reflected energy; after that point, the two remain approximately equal.
- The cardioid is more susceptible to the problems of "P-pop" (the blast of explosives such as "P" and "T" in speech), wind noise, and handling or mechanical noise.
- The cardioid is more complex to design and construct. Therefore, expect to pay more, sometimes, for a cardioid microphone than for an omni of seemingly equal audio quality. The cardioid's more complex construction results in the omni often being the more rugged of the two as well.
- The cardioid offers greater resistance to feedback in most sound reinforcement applications. This results from its lower REE and is further aided by the proximity effect.
- The cardioid increases the effective working distance. The Figure 3.3-7 comparison chart lists a *distance factor* (DF) of 1.7 for the cardioid, meaning that a cardioid microphone has a working distance advantage over the omni of 1.7:1. This is calculated on the assumption of a perfect cardioid, in a totally diffuse noise field. An ideal cardioid, then, could be used at a distance of 1.7 times that of the omni for a given ratio of desired, on-axis signal to ambient noise.

In the real world, nonlinear polar response and the inability of cardioid microphones to achieve total cancellation at their null would seem to reduce greatly the cardioid's working distance advantage. It should be noted that in actual use, however, increasing the working distance often has more to do with attenuating a single, offending noise source than with overcoming a diffuse noise field. Directing the deep null of a good cardioid microphone at a whirring still-camera motor drive at a press conference may offer more than a 1.7:1 working distance advantage over an omni.

The cardioid exhibits proximity effect. While some designs are quite low in proximity effect, all exhibit some bass-boost phenomenon when used close. Although this may be considered an enhancement in many close-miking applications, care should be taken to avoid input overload or loss of intelligibility that may result from excessive proximity effect.

Other Patterns

An examination of Figure 3.3-7 will show how the three other polar patterns compare to the omni and the cardioid. The hypercardioid, for example, combines a tight acceptance angle with superior side rejection and offers the lowest REE. The result is a good pattern for distant miking (often required on a boom mount). Most short shotgun microphones approximate a hypercardioid pattern. Notice that the bidirectional pattern offers the best side rejection, but with no advantage over the cardioid in REE. Bidirectional microphones are typically ribbons or dual-diaphragm condensers. They are excellent for picking up two sound sources in music or dialogue miking with no phase problems. Compared to the omni, each of the directional patterns exhibits all six of the characteristics described previously.

Working Distance

Sometimes it is not sufficient to reduce ambient noise merely by using a polar pattern that offers the lowest REE. In very noisy environments (for example, in an aircraft, on a battlefield, in a factory, or at a sporting event) it may be desirable to differentiate between close sound (an announcer, for example) and distant sound. Microphones that offer considerable proximity effect may be used to advantage in these situations. Close-mic the talent and roll-off the low end as needed to flatten the response. In extreme situations, a noise-canceling (differential) microphone may be required.

Because of the special design (rear ports and back damping systems) of the differential microphone, sound arriving from a distance strikes both sides of the diaphragm with equal intensity and in phase. A positive pressure on the front, for example, would encounter a positive pressure on the rear of the diaphragm, causing the signal to be canceled. A combination of inverse square law and port damping causes sound that originates very close to the front of the noise-canceling microphone to be far lower in intensity and to exhibit some phase error by the time it arrives at the rear of the diaphragm. The result is diaphragm movement and maximum output. The noise-canceling microphone is able, therefore, to differentiate between close and distant sound sources. The audio quality of such systems normally limits them to voice communication applications.

Inverse Square Law

The easiest, and certainly the least expensive way to limit the apparent working distance of a microphone is by positioning the microphone very close to the sound source. Inverse square law shows that a decrease in the distance between the microphone and the sound

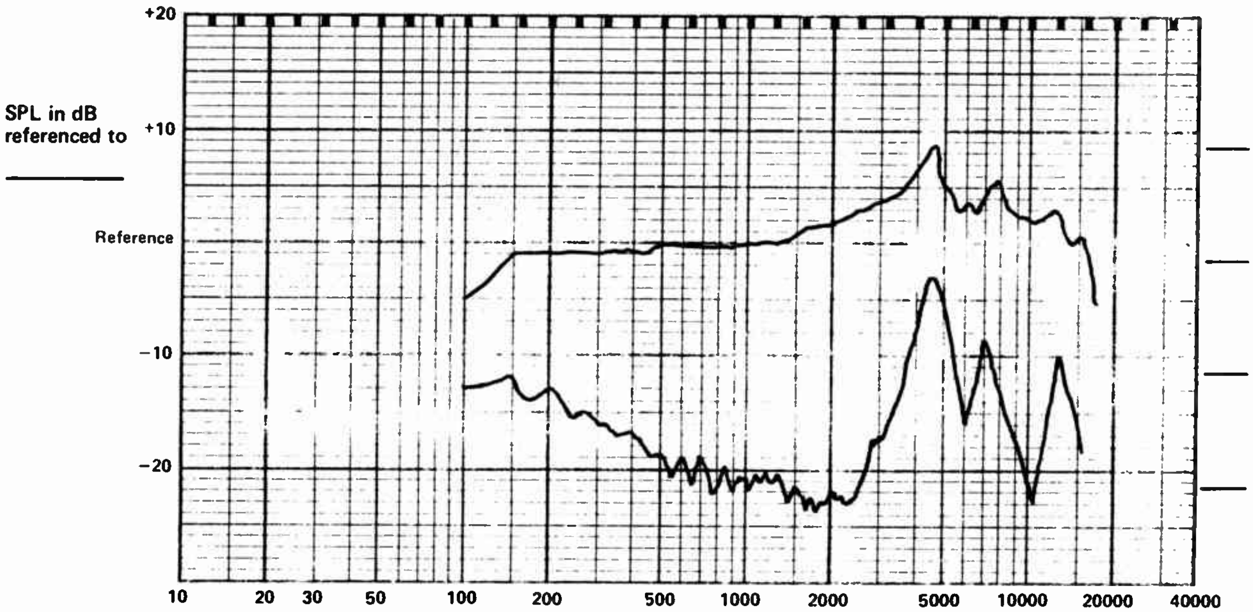


Figure 3.3-11(a). Front and back curves of a typical cardioid hand-held vocal microphone.

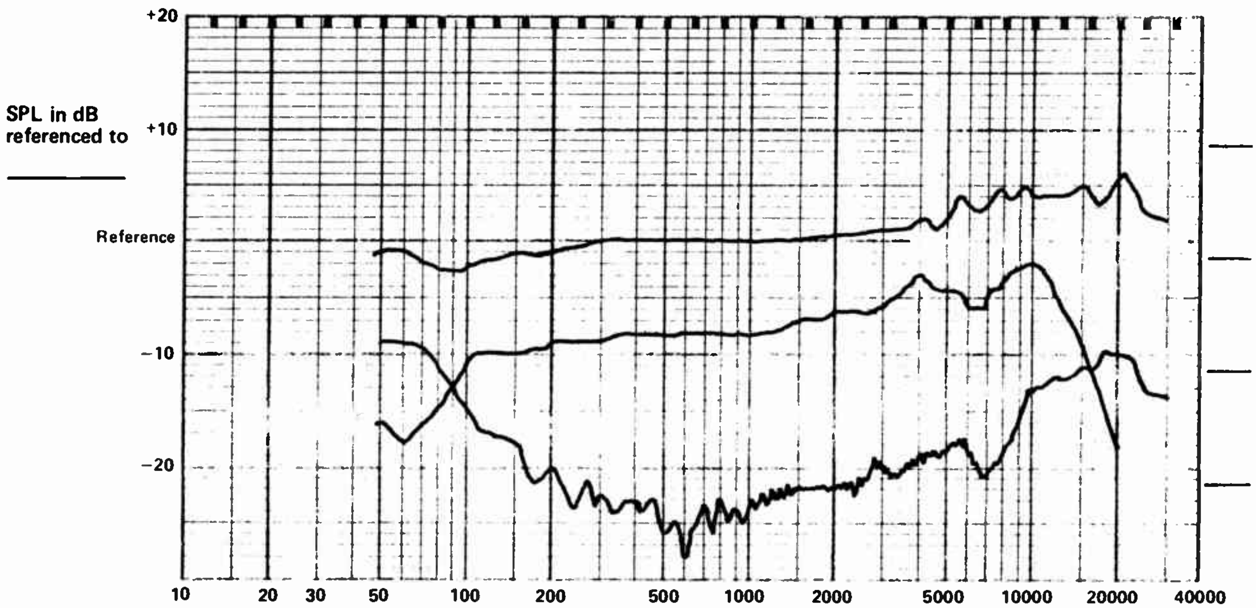


Figure 3.3-11(b). Front, side and back curves of a high- quality cardioid condenser microphone.

source at 90° off-axis. Here, a properly designed cardioid emulates well its mathematical model. At 180° off-axis, however, the cardioid can not live up to the equation. The chart indicates 0 output. In reality, well designed cardioids are capable of significant cancellation at 180° but, typically, only on the order of 20 dB. Still, a 20 dB drop in the level of some unwanted noise may be enjoyed simply by turning the cardioid to face directly away from the noise. That is equivalent to moving the sound source to ten times its actual distance from the microphone.

The 180° response curves (back curve) of many cardioid microphones show their tendency to more closely resemble omnis at both the low and high frequencies. The much more impressive cancellation in the midrange offers just the kind of temptation that sometimes causes a manufacturer to release a data sheet that shows one polar pattern only, and that at some unknown frequency. One of the most beneficial performance advantages an end-user should look for in a well-designed microphone is off-axis linearity, as illustrated in Figures 3.3-11 (a) and (b).

First, check to see whether the scaling is logarithmic or linear. A log scale (the most commonly used) will show a fairly modest inward curve of the cardioid pattern at 90°, indicating a 6 dB drop in level. The linear scale polar for the same microphone will show a polar pattern that appears much more directional. The outside circle of the linear polar represents 100% while the center of the circle equals 0 output. Because a 6 dB loss is equal to a 50% drop in voltage, the polar curve at 90° sweeps in to half the distance between the outside of the circle and its center. Second, determine the graduations between concentric circles. Are the darker lines 5 dB or 10 dB apart? Finally, take note of the dynamic range of the polar pattern. This may be determined by counting graduation lines inward from the point where the polar crosses 0°, in 5 or 10 dB steps (as marked) to the smallest inner circle. Polars may be found in most any range, with 25, 30 and 40 dB all being common. These differences will also alter the shape of a polar pattern.

Keep in mind that the polar pattern represents a cross-sectional, two dimensional diagram of a three dimensional function. The 131° arc, for example, that is described by the 3 dB down points on either side of the axis of the cardioid microphone can really best be thought of as a conical area within which the microphone is virtually uniformly sensitive. This area is often referred to as the microphone's *angle of acceptance or included angle.*

Omni Observations

The omnidirectional microphone is the easiest type to make. It consists of a diaphragm and generating element backed by a totally sealed case. When placed in a sound pressure field, the perfect omni disregards the direction of the sound's origin. A positive pressure (air expanding) at the diaphragm, for example, causes the diaphragm to move inward regardless of the sound's point of origin (see Figure 3.3-9). Such a microphone may also be referred to as a *pressure microphone.*

So it would be with the perfect omni. Most omnidirectional microphones, however, are not truly omnidirectional. The case of the microphone represents a barrier to higher frequencies arriving from off-axis. Because of this case effect, most omnis are increasingly directional at higher frequencies. The smaller the omni, the more truly omnidirectional it may be. In addition to the case effect, energy arriving at the diaphragm from on-axis is reinforced at those frequencies to which the size of this *baffle area* is significant. The baffle effect causes a rise in the microphone's high-frequency output, but only with respect to energy arriving on axis.

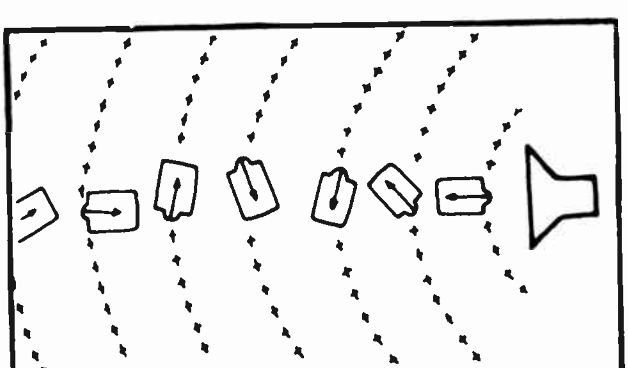


Figure 3.3-9. Omnidirectional microphone principle.

Cardioid Considerations

Directional microphones employ a damped porting system in their element design that allows sound waves to act upon the rear of the diaphragm as well as the front. The design introduces varying amounts of phase shift for sound arriving from off-axis, resulting in cancellation. The rear entry ports of most directional microphones are spaced at a single distance or "D" from the diaphragm (see Figure 3.3-10). Multiple port systems are also available and are designed to reduce the proximity effect.

Care should be taken when using any directional microphone not to obstruct the ports with the microphone clip, hand, clothing, gaffer's tape or logo flags. Covering even some of the ports may result in serious degradation of the microphone's directional characteristics and overall sound quality. The chart given in Figure 3.3-7 shows how the various patterns should relate to the reference omni in their ability to reject unwanted energy arriving from various points off-axis. A sound source that delivers 60 dB sound pressure level (SPL) to a cardioid on-axis from one foot away will sound to the microphone as if it has dropped in level to 54 dB (or been moved to two feet away) when the microphone is simply rotated to position the sound

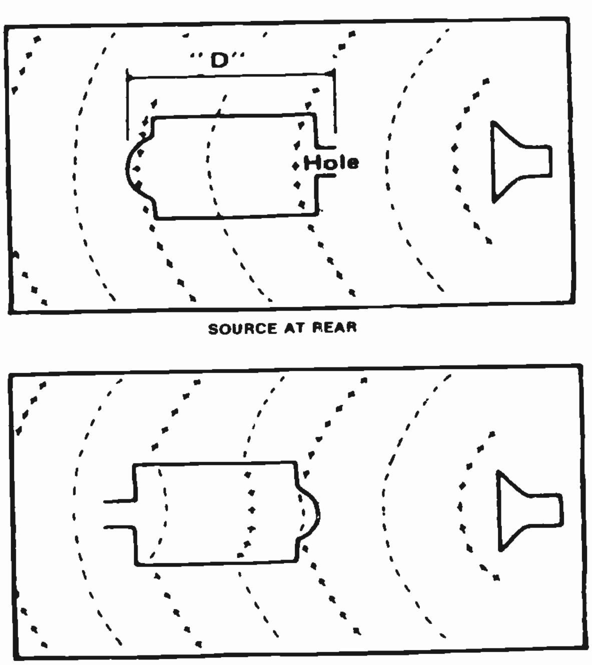
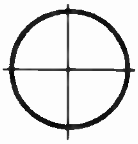
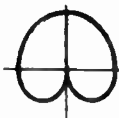
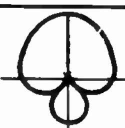
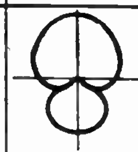
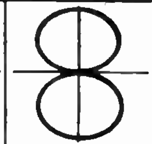


Figure 3.3-10. Single-D cardioid microphone operating principle.

CHARACTERISTIC	OMNI-DIRECTIONAL	CARDIOID	SUPER-CARDIOID	HYPER-CARDIOID	BIDIRECTIONAL
Polar response pattern					
Polar equation	1	.5 + .5 cos θ	.375 + .625 cos θ	.25 + .75 cos θ	cos θ
Pickup ARC 3 dB down (1)	—	131°	115°	105°	90°
Pickup ARC 6dB down	—	180°	156°	141°	120°
Relative output at 90° dB	0	−6	−8.6	−12	−∞
Relative output at 180° dB	0	−∞	−11.7	−6	0
Angle at which output = 0	—	180°	126°	110°	90°
Random energy efficiency (REE)	1 0dB	.333 −4.8dB	.268 −5.7dB(2)	.250 −6.0dB(3)	.333 −4.8dB
Distance factor (DF)	1	1.7	1.9	2	1.7

NOTE:

1 = Drawn shaded on polar pattern
2 = Maximum front-to-total random energy efficiency for a first order cardioid
3 = Minimum random energy efficiency for a first order cardioid

Figure 3.3-7. Microphone polar patterns and their characteristics.

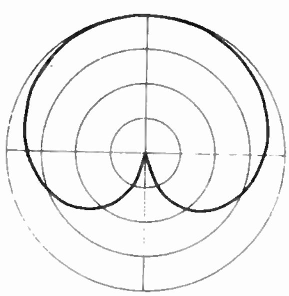


Figure 3.3-8(a). Cardioid log scale polar. Scale is 10 dB per division.

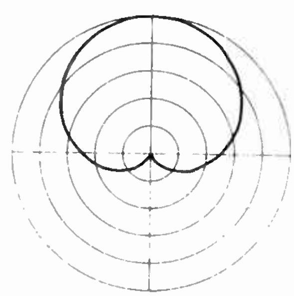


Figure 3.3-8(b). Cardioid linear scale polar.

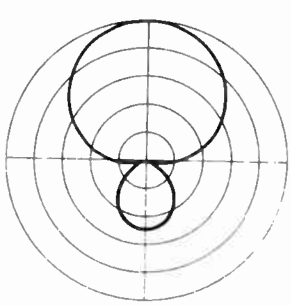


Figure 3.3-8(c). Hypercardioid linear scale polar. Dynamic range is 50 dB.

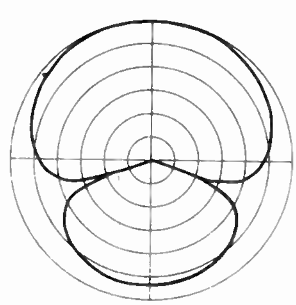


Figure 3.3-8(d). Hypercardioid log scale polar. Scale is 10 dB per division. Dynamic range is 40 dB.

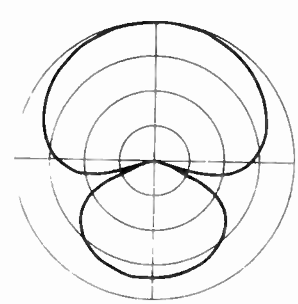


Figure 3.3-8(e). Hypercardioid log scale polar. Scale is 10 dB per division. Dynamic range is 60 dB.

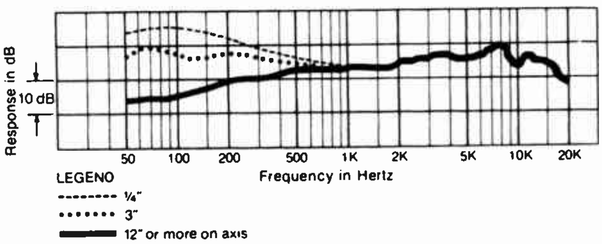


Figure 3.3-5. Influence of proximity effect on a directional microphone response. (Courtesy of Audio-Technica U.S., Inc.)

amount of proximity effect that is possible to attain. The response curve of a particular directional microphone, tested at a specific (or perhaps even, unknown) distance may be of little value to someone who wishes to use the microphone at another distance. Ideally, directional microphone data should include close and distant curves, as illustrated in Figure 3.3-5.

Polar Patterns

As difficult as it might seem at times, to pick up a desired audio signal the real problem more often lies in eliminating unwanted sounds. Microphones with various directional patterns are often used to improve the ratio of desired signal to ambient noise or other unwanted sounds.

For most miking applications, a suitable on-axis curve alone may not be satisfactory. Ambient noise, leakage from other instruments in a band or orchestra, room reverberation and feedback potential from PA floor monitors are some of the reasons why it is important to know the off-axis response of a microphone.

The best view of the microphone's off-axis response is obtained by examining several different polar plots. These should be drawn at low, mid-band, and high frequencies. Overlaid, these plots should reveal how well the microphone maintains its directionality at each frequency (illustrated in Figure 3.3-6).

There are several broad classes based on fundamental polar patterns, to which most microphones directional characteristics conform, to some extent or another. These include:

Omnidirectional

- Bidirectional
- Cardioid
- Supercardioid
- Hypercardioid

Figure 3.3-7 shows the relative data for each of the patterns. The data, however, is taken from mathematical models representing the perfect polar characteristic for each example. Actual microphone polars may vary from near perfection to close resemblance. In the real world, the microphone design engineer must go beyond math equations to accomplish a desired axial response and sensitivity while maintaining polar uniformity. It is truly a blending of art and science.

Polar Scaling and Dynamic Range

Care should be taken when reading polar patterns to observe several variables in the way that they are represented. These variables can drastically alter one's perception of the microphone's directionality if they are not examined closely (see Figures 3.3-8 a-e).

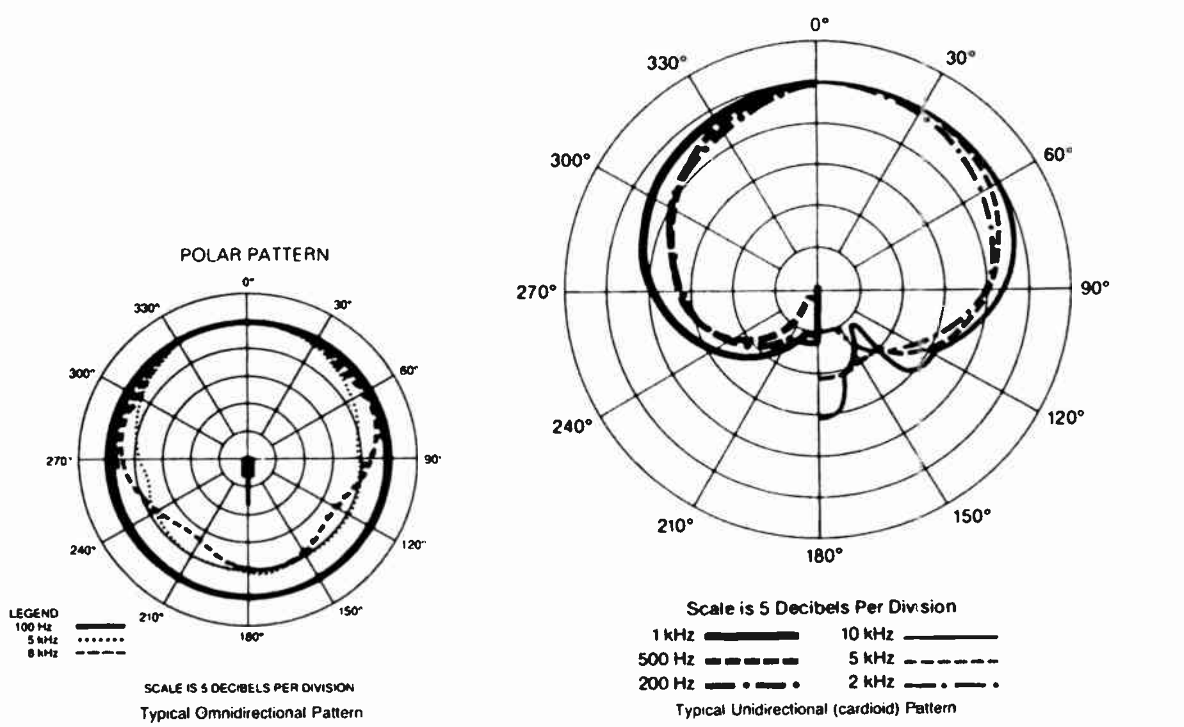


Figure 3.3-6. Polar patterns drawn at several frequencies. (Courtesy of Audio-Technica U.S., Inc.)

of these supplies, some modern discrete condenser designs require 48 V. The phantom supply voltage in nonelectret condenser microphones is often stepped-up by an internal circuit to provide a sufficient capacitor-polarizing charge for good signal-to-noise figures.

The amount of current delivered by the power supply is also a factor. While some condenser microphones can operate with less than a milliamp, others require up to 2.5 mA or more. When insufficient current is provided, the level of the microphone is reduced and the signal is distorted.

In a phantom power circuit, the plus side of the dc supply is applied equally to both of the signal conducting leads of a balanced microphone line. This may be done by means of either a build-out of matched precision resistors, or via a center-tapped transformer. In each case, the return path is the shield. In the microphone, the dc may be similarly tapped via the resistor or center-tapped transformer method to provide the power it needs. The dc is prevented from appearing at the impedance converter output by dc-blocking capacitors or the internal center-tapped transformer.

If a balanced-output dynamic microphone is connected to a line with phantom power present, performance should not be altered, nor should damage occur to the dynamic element. The voice coil or output transformer winding connects across the two signal leads and should see no potential difference between them. Because there is no connection between either lead and the shield (the dc return path), there is no circuit. If an unbalanced dynamic microphone is connected to a phantom supply, the dc will pass through the voice coil and probably destroy it. A less common powering system called "A/B" or "T" powering is not compatible with dynamic or phantom-powered microphones. A/B power puts the positive side of the dc on one signal lead and the negative on the other. This will damage even a balanced output dynamic microphone.

Microphone specifications

Charts, graphs and specifications are usually available for a professional microphone. Some microphones come with individual graphs for frequency response and polar pattern. While specs tell part of the story, there is nothing better than trying a microphone in a specific application. Be aware that the microphone preamp and microphone cable can both contribute to the sound of the microphone. The same microphone may sound different when amplified through another preamp.

Frequency Response

One of the first specifications considered on a microphone data sheet is the frequency response range or limits. The required low and high-frequency limits may depend upon the nature of the sound source that is being miked, the medium by which the signal is to be stored and transmitted and the environment in which the miking is to be done. Often, unfortunately, more attention is given to the response limits than to how the microphone actually sounds in its intended application. Nonlinearities of the response often contribute more to the listener's subjective impression of sound transmitted by the microphone than do the response limits. A specification that reads "frequency response 40 to 18,000 Hz," by itself, says little about a microphone's actual sound in use. Add to that some limits, like ± 3.0 dB, and one's knowledge of the microphone, while improved, is still lacking.

The shape of the response and polar response characteristics will tend to contribute to the character or personality of the microphone's sound. Response nonlinearities can create acoustic feedback in sound reinforcement, nasality, poor intelligibility, excessive sibilance, muffled sound or any of a variety of other acoustic problems. On the other hand, a microphone's response may be deliberately, carefully tailored by the design engineer to solve problems rather than create them. A rolled-off low frequency response and a rising high frequency response may be employed in a microphone that is intended for use at a considerable distance. This tailoring can reduce the effects of unwanted low frequency information, such as traffic noise or the rumble of air handling systems, while boosting the high frequencies that are normally attenuated with distance.

Some microphones that are intended to be worn on the body exhibit a response that compensates for the chest cavity resonance which they tend to pick up in a lavaliere mounting position. A rolled-off low end response may help considerably in attenuating handling or stand-born noise as well as wind noise or the breath blasts of plosives in speech.

Unfortunately, published specifications can serve as only one guide in understanding a microphone. The extent to which they serve any function at all is, of course, dependent upon the credibility of the source. This may help to explain why a very poor sounding \$50 microphone might show limited specifications that appear superior to an excellent sounding \$250 microphone. Certainly, a response curve drawn with the microphone directly facing the sound source (on axis response) will give us a clearer picture of the instrument than will a statement of its response limits. Even curves from well-respected manufacturers, however, may be difficult to compare due to the variety of test procedures and standards. For example, the frequency response chart is an X-Y graph that compares decibel output to frequency. If that chart is compressed vertically or stretched horizontally, the response curve will automatically appear more linear. Frequency response plots for the same microphone may also vary greatly when run on the same equipment, at the same scale. Such variables as the recording speed, damping or even the direction of the tone sweep (low to high or high to low) may result in vastly different curves.

Directional microphones present still another possible anomaly in the testing process. These microphones exhibit a phenomenon known as the *proximity effect,* which results in a bass-boosted output when close-miking a small sound source. Proximity effect is neither good nor bad; its value is dependent upon the intended application. Various designs will differ in the

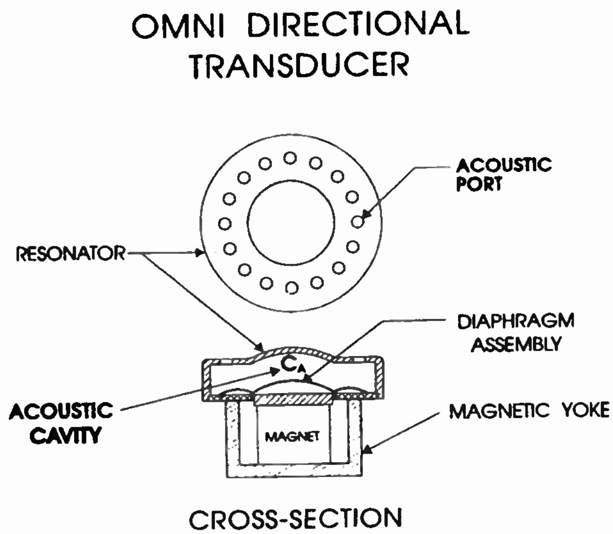


Figure 3.3-2. Cutaway drawing of a dynamic element showing helmholtz resonators. (Courtesy of Electro-Voice, Inc.)

transformer within their housing to correct for design trade-offs in their voice coil. The transformer adds to the microphone's cost and may also restrict performance if it is not a high quality unit, limiting the frequency response and possibly increasing distortion. But properly designed dynamic microphones can be the most rugged of the high-quality transducer types; some have truly become legendary for their ability to provide high quality broadcast audio with virtual bullet-proof construction.

Condenser Microphones

In the condenser microphone, a capacitor forms the generating element, as illustrated in Figure 3.3-3. One side of the capacitor is the diaphragm, the other is the fixed backplate. Air between these two plates acts as a dielectric. The capacitor, of course, must possess a positive electrical charge on one plate and a negative charge on the other. The conventional or discrete condenser receives this polarizing or bias voltage from an external dc power supply. In older condenser microphones, separate leads were required to deliver dc from the power supply to the microphone. Today, *phantom* power is used in most conventional condenser systems to deliver the required dc voltage to the microphone

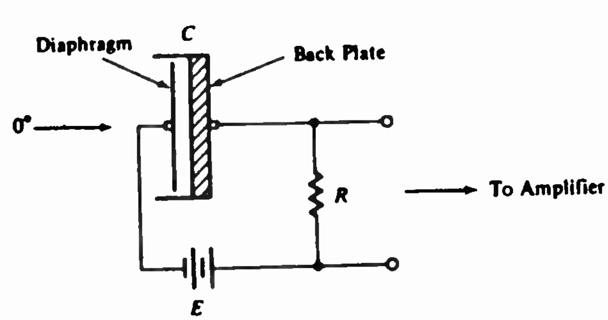


Figure 3.3-3. Conventional capacitor microphone system.

over the same conductors and their shield that are used to carry the audio signal.

Upon activation of the power supply, a positive voltage is quickly built up on the rear surface of the diaphragm. This causes an electrical current to flow through the resistor until the surface of the backplate finally receives a negative charge of equal value. As sound waves (air pressure changes) strike the diaphragm, causing it to move back and forth, the distance between the two plates rapidly increases and decreases. This causes proportional changes in the capacity of the condenser to hold a charge. The result is an ac current flow in the resistor and a voltage across the resistor that corresponds to the excursion of the diaphragm. While this voltage effectively represents the output voltage of the microphone, the source impedance is far too high to be carried for any distance over microphone cable. This output signal, then, is presented to an impedance converter circuit, usually a vacuum tube or FET inside the microphone. Power for the impedance converter is derived from the same source that provides the polarizing voltage for the element. The impedance converter delivers a low impedance output that can be fed down long microphone cables with minimal loss.

Electret Condenser

The electret condenser microphone utilizes a material which has the ability to hold a charge applied during the manufacturing process. Most high-quality electrets apply this material to the fixed back-plate of the capacitor, as shown in Figure 3.3-4. Some designs employ a charged diaphragm instead, but pay several performance penalties in doing so. Lowering the weight of the diaphragm by moving the electret material to the backplate results in lower handling noise, extended frequency response and improved transient response.

Although the electret functions much like the discrete condenser, but produces its output voltage without the need for an external high-voltage dc supply, an impedance converter is still required. The low voltage needed to power it, however, may be derived from internal or external batteries or an external ac-powered supply.

Phantom Power For Condenser Microphones

Phantom power, or *simplex power,* provides one means for remotely powering condenser microphones. Phantom power requirements of various microphones may fall anywhere between 9 and 48 V. External supplies are most often designed to deliver 9, 12, 18, 24, 30, or 48 V. While many electrets will operate on any

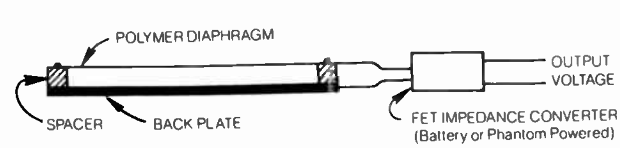


Figure 3.3-4. Electret condenser element. (Courtesy of Audio-Technica U.S., Inc.)

sympathetically-induced mechanical recreation of the amplitude and frequency of the sound presented to it.

The metallic ribbon's movement across the magnet's flux lines induces a small ac voltage onto the ribbon. The leads from the ribbon's ends connect it to a step-up transformer which converts the extremely low impedance of the ribbon (approximately 1 Ω) to a usable figure which might lie between 50 and 500 Ω. Ribbon microphones are available in a variety of fixed or variable patterns. They are known for delivering a very warm sound, due to the ribbon's sensitivity to low frequencies. They are also known for their fragility. Blowing into a ribbon microphone can destroy the ribbon. Some manufacturers suggest that ribbon microphones be stored in positions so as not to let the ribbon sag. Even rapid panning on a studio boom has caused ribbon failure. Newer designs, however, have provided considerable improvements in durability and a lower failure rate.

Some ribbon microphones should not be connected to a microphone preamp with active phantom or A/B powering. The voltage from these supplies can damage the ribbon.

Dynamic Moving Coil Microphones

Although the ribbon microphone is a type of dynamic microphone, in common usage, the term dynamic microphone usually refers to a microphone with a moving coil. The dynamic microphone has a diaphragm with its back attached to a voice coil, as illustrated in Figure 3.3-1. This extremely lightweight coil of wire is suspended in a magnetic field supplied by a permanent magnet structure. The ends of the voice coil are brought out to stronger leads, which connect either to a transformer or the microphone's output connector.

Sound waves reaching the diaphragm cause it to move back and forth in sympathy. The attached voice coil is then forced to cut the lines of flux in the magnetic field causing a small ac voltage to appear at the ends of the leads. This signal closely emulates the sound

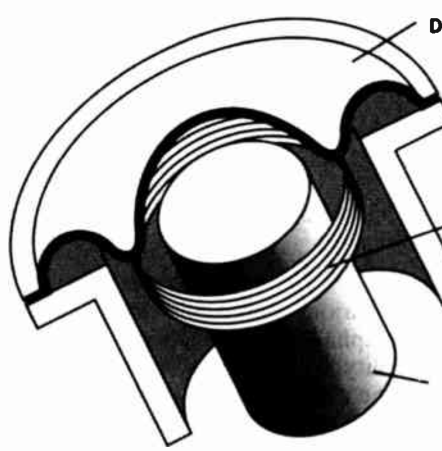


Figure 3.3-1. Dynamic moving coil element. (Courtesy of Shure Brothers Incorporated.)

waves in frequency and amplitude. The diaphragm design is crucial to good dynamic microphone performance. It must be highly compliant to allow effortless excursion at all frequencies of interest. In addition, this movement must be accomplished with maximum linearity and a minimum of break-up modes. Nonlinearities of diaphragm movement would include, for example, any tendency for a rocking motion to be set up which, at some frequencies, would allow one side of the voice coil to be travelling downward while the other side is moving upward. Such rocking results in phase cancellation and a dip in the frequency response. Break-up modes occur when a portion of the diaphragm resonates independently of the rest of the surface. Again, phase cancellation results and, with it, response anomalies occur.

The design and construction of a high quality dynamic microphone suitable for broadcasting use is a blending of science and art. As with other transducers, much of its design may be modeled by the use of equivalent electrical circuits. The process is similar to the computer-assisted design of loudspeakers using Thiel parameters. In the microphone, however, the seemingly infinite variety of complexities in certain parts of its design is so great that the artistry of the microphone engineer is constantly called upon. In the design process, the goals set before the microphone design engineer are sometimes all but impossible to achieve within the same product. As is true in other areas of engineering, design trade-offs are numerous and the laws of physics tend to win in the end.

Size plays an important role in the performance of the dynamic microphone. Small dynamic mechanisms tend to be very inefficient. Their acoustic sensitivity is low, while their mechanical sensitivity is high by comparison. The result may be a poor system signal-to-noise ratio (SNR) and a lot of handling or noise transmitted through the microphone stand. Internal shock-mount systems may be used to reduce the mechanical excitation but the design goal of small size may then be defeated. Small size usually means sacrificed low-frequency response in dynamic microphones. This is not to say that large dynamic microphones will always have an extended low frequency response. Very small dynamics will almost certainly, though, be lacking in low frequency output. Another physical characteristic of the dynamic microphone that affects its performance is the mass of the diaphragm/voice coil assembly. The greater the mass, the more limited will be the high frequency response. Common design practice includes the use of Helmholtz resonators immediately in front of the diaphragm. These create peaks and effectively extend high frequency response beyond the normal limits of the system (see Figure 3.3-2). Advances in metallurgy, specifically the use of stronger Neodymium magnets, has have resulted in dynamic microphones with higher outputs.

Most broadcast-quality dynamic low-Z microphones exhibit an impedance that is a function of the turns and gauge of their voice coil wire. Some older, more public address-oriented microphones employ a than directional microphones, however, some small-diaphragm condenser omni microphones are quite sensitive to popping.

Microphones with directional patterns, cardioid (heart-shaped) or figure-eight, are used in broadcasting to keep ambient noises from being picked up. Cardioid microphones are available by degree of directionality—cardioid, hypercardioid, supercardioid. Shotgun microphones, sometimes referred to as supercardioids, offer superior directionality. While their "reach," or ability to hear sounds at a distance is greater than hypercardioid patterns, shotgun microphones are omnidirectional at middle to lower frequencies. They are best used on a large well-damped sound stage or out in the open.

Figure-eight patterned microphones are useful when placed directly between two sound sources. Even so, cardioids are chosen over figure-eights in broadcasting because exact placement can be difficult and because two cardioids can be set at different levels to accommodate sound sources with different levels. Due to the physics of their design, cardioid microphones usually exhibit more bass response than omnis when the distance between microphone and person speaking is one foot or less. This phenomenon is called the *proximity effect*. If not abused, the proximity effect adds warmth and power to most voices. Directional microphones are also typically found in broadcast studios because they can be positioned to minimize noise from tape recorders, wire copy machines HVAC ducts, computer drives and fans and other noise generating devices.

Output Level

Output level is important because all microphone preamps add a certain amount of noise, especially at the upper end of their operating range. The higher the output of the microphone, the less amplification will be required from the microphone preamp. In general, condenser microphones have higher outputs than dynamic and ribbon microphones, although the use of higher flux magnets in both dynamic and ribbon microphones will increase their output.

Ruggedness is more important in ENG and electronic field production (EFP) recording. Microphones for these applications need to be able to withstand rough handling and harsh elements. Ribbon microphones are seldom used because of the fragility of the ribbon element. Some condenser microphones are better suited for highly controlled environments because they are sensitive to humidity, temperature changes and wind. Other condenser microphones are more tolerant of these conditions.

Hot Spot

The hot spot (sometimes called a sweet spot) of a microphone is that area within its pickup pattern in which a particular sound source sounds best. It is not an absolute position, but a relative one. In a very quiet and well-damped environment, the hot spot may be larger. However, if the sound source is thin and would benefit from more bass response, the hot spot for a cardioid microphone might be limited to an area within the range of its proximity effect. If the sound source has an overabundance of low frequencies, the area providing the proximity effect should be avoided.

If the sound source, an announcer for example, moves around too much, a hypercardioid pattern may provide too narrow a hot spot, resulting in noticeably uneven levels and sense of presence. If the announcer can not be trained to stay within the hotspot, a more forgiving cardioid pattern may be a better choice.

Regardless of the style or pattern of microphones chosen, the best results require the use of a high-quality, well-shielded microphone cable. Quality cable is more important here than anywhere in the audio chain because of the high amount of gain applied by the microphone preamplifier. The preamplifier raises the relatively low level of the microphone to line level. Poorly shielded, or low quality cable may not be able to reject electromagnetic or radio frequency induced noise. As a result, that low-level noise is amplified by the microphone preamplifier and can become audible. It is a good idea to start with well-shielded cables in which the shield is grounded at both ends to the XLR connectors. Using these cables to interconnect other pieces may result in ground loops, which can cause hums and buzzes. The most conservative solution is to keep microphone cables separate from those used to connect active, powered circuits.

MICROPHONES IN DETAIL

Microphones will be described in terms of their generating element (transducer) types:

- Ribbon
- Dynamic
- Moving coil
- Condenser

Transducer Types

A microphone is a transducer: a device which, when activated by energy from one system, converts or transduces that energy to another form. In the microphone, acoustical energy (sound waves impinging on the diaphragm) is converted to a varying voltage that is the electrical analog of the sound. The method by which the microphone converts acoustical energy to electrical energy is one of the ways by which microphones are differentiated.

Dynamic Ribbon Microphones

The ribbon, or velocity microphone, utilizes a very thin, corrugated metallic foil ribbon suspended within the flux field of a permanent magnet. The ribbon microphone is also referred to as a *velocity* microphone because it responds to air particle velocity at the ribbon. The term *pressure gradient* is also used to denote that the pressure exerted on the ribbon is in proportion to the difference between the pressures present on each side. While the ribbon's ends are held in place, the rest of it is allowed to move freely back and forth in a

3.3 MICROPHONES

TY FORD
TECHNIQUE, INC., BALTIMORE, MD
AND GREG SILSBY

INTRODUCTION

Improving a broadcast facility's sound by choosing the right microphone is the direct result of knowing enough about microphones to make the right decisions. When considering both budget and application, there are usually a number of correct choices. This chapter is intended to provide you with information that will help you make the right choices. To that end, this chapter reviews the basic types of microphones, their construction, and how their different characteristics can be used in various applications for best results.

The successful use of any microphone also depends on the acoustic environment in which it is used and the particular voice or instrument on which the microphone is used. The problems of poor acoustic environments can not always be overcome by using more expensive microphones. Instead of great results, you might end up with technically excellent audio and a great performance in an obviously bad acoustical environment. In addition, individual voices vary greatly. The frequency response and texture of some voices seem to lend themselves to the electronic medium. Other voices with quirky frequency response, sibilance and other artifacts can be problematic, even after considerable processing. Great sounding audio, then, is the result of a great source, a great environment and the right microphone.

Of the many microphones available, broadcasters usually choose hand-held, boom-mounted and lavaliere styles. The hand held microphone is used for on-camera and electronic newsgathering (ENG). Boom-mounted microphones are usually found in the broadcast studio and some hand-held microphones also are boom-mounted. The lavaliere microphone is a small clip-on usually used in TV news, where a hand-held or boom-mounted microphone would be obtrusive. Lavaliere microphones can also be used in radio talk show programming where the guest may be unfamiliar with good microphone technique or intimidated by a large boom-mounted microphone.

The four attributes of a microphone most broadcasters need to consider are self-noise, directionality, output level, and ruggedness.

Self-Noise

Self-noise is normally perceived as a hiss. It is the electronic noise generated by the active circuitry of the microphone itself. Condenser microphones use a vacuum tube or field effect transistor (FET) circuitry to convert the high impedance of the condenser capsule to a lower impedance, more suited for the input of a microphone preamp. Vacuum tube condenser microphones are generally noisier than condenser microphones that use FETs. However, careful selection of the vacuum tube can result in a tube microphone being quieter than its FET counterpart.

Self-noise is an important factor in studio recordings, where every attempt to eliminate extraneous noise is important. Self-noise is less important in applications where the ambient noise level is considerably higher. The ambient noise level in a broadcast air studio may be quite low, equaling that of a well-designed music studio. It may also be rather high due to noise from HVAC systems, tape machines and computer hard drives and fans, wire copy machines and sound transferred through the structure from the surrounding environment. If the combined noise from any or all of these sources can not be reduced, spending extra money for a microphone with exceptionally self noise is not the best use of money. Finding the best sounding, most directional microphone and positioning it to reject as many of the ambient noises as possible is a better approach.

The self noise of dynamic or ribbon microphones is usually much less because these microphones do not use active electronics. However, the output of dynamic and ribbon microphones is usually lower than that of a condenser. Because of this, more gain is required of the microphone preamplifier to bring the dynamic and ribbon microphones up to the appropriate operating level. Less expensive or poorly designed microphone preamps also generate noise, usually perceived as hiss.

Directionality

Omnidirectional and directional microphones each have their place in broadcasting. Omnidirectional microphones pick up most sounds from all directions equally. However, they are usually directional at high frequencies. Omnidirectional microphones are usually used in news gathering interviews in which one or more voices are intended to be picked up. Their wide pattern is useful because the position of the microphone does not have to be moved from person to person. On the other hand, if there is a lot of ambient noise, or if only one voice is needed, a directional microphone may be more effective. In general, omnidirectional microphones are also usually less prone to "popping" achieved when the supply or return duct paths connecting two rooms are longer. The optimum situation is to have completely separate ducts back to the fan source; however, this may not be practical in some situations.

CONCLUSIONS

This chapter has presented basic principles of acoustic design that are relevant to the design or renovation of broadcast facilities. Hopefully, this material will offer some perspective and insight to the reader. There are, however, many different techniques of using these principles to achieve specific acoustic design goals and a wealth of literature and experience exists in this area. The value of a professional acoustical consultant in a major design or renovation project should not be underestimated. Acoustical design techniques can be very expensive and mistakes can be even costlier. With a basic understanding of acoustic principles, users will be better able to explain their needs to a consultant, understand the ramifications of various design trade-offs and develop the ability to make informed and competent decisions in the area of acoustics.

BIBLIOGRAPHY

Backus, J., The Acoustical Foundations of Music, W. Norton, 1977.

Ballou, G., *Handbook for Sound Engineers: The New Audio Cyclopedia, 2nd Ed.*, Focal Press, 1991.

Beranek, L. L., *Acoustics*, Acoustical Soc. of America, 1986.

Beranek, L. L., *Acoustic Measurements*, Acoustical Soc. of America, 1988.

Beranek, L. L, *Music, Acoustics and Architecture*, John Wiley, 1962. (out-of-print)

Beranek, L. L., & I. L. Vér: *Noise and Vibration Control Engineering*, John Wiley, 1992.

Cooper, J., *Building a Recording Studio*, Recording Inst. of America, 1978

Crocker, M. J., *Handbook of Acoustics*, John Wiley & Sons, 1998.

Davis, D., *Acoustical Tests and Measurements*, H. Sams, 1965. (o-p)

Davis, D., *Sound System Engineering, 2nd Ed.*, H. Sams, 1987. (o-p)

Doelle, L., *Environmental Acoustics*, McGraw-Hill, 1972. (o-p)

Egan, M. D., *Architectural Acoustics*, McGraw-Hill, 1988

Everest, F. A., *Acoustic Techniques for Home and Studio, 2nd Ed.*, Tab, 1984. (o-p)

Everest, F. A., *How to Build a Small Budget Recording Studio from Scratch . . . With 12 Tested Designs, 2nd Ed.*, Tab, 1988. (o-p)

Everest, F. A., *The Master Handbook of Acoustics, 3rd Ed.*, Tab, 1994.

Everest, F. A., *Sound Studio Construction on a Budget*, McGraw-Hill, 1996.

Hall, D. E., *Basic Acoustics*, Krieger, 1992.

Knudsen, V. and C. Harris, *Acoustical Designing in Architecture*, Acoustical Soc. of America, 1978.

Mankovsky, V., Acoustics of Studios and Auditoria, Hastings House, 1971. (o-p)

Rettinger, M., *Acoustical Design and Noise Control, Vol. 1 & 2.*, Chemical Publ. Co., 1977. (o-p)

Rettinger, M., *Handbook of Architectural Acoustics and Noise Control*, Tab, 1998.

Rettinger, M., *Studio Acoustics*, Chemical Pub. Co., 1981. (o-p).

Yerges, L. F., *Sound, Noise and Vibration Control*, Krieger, 1978.

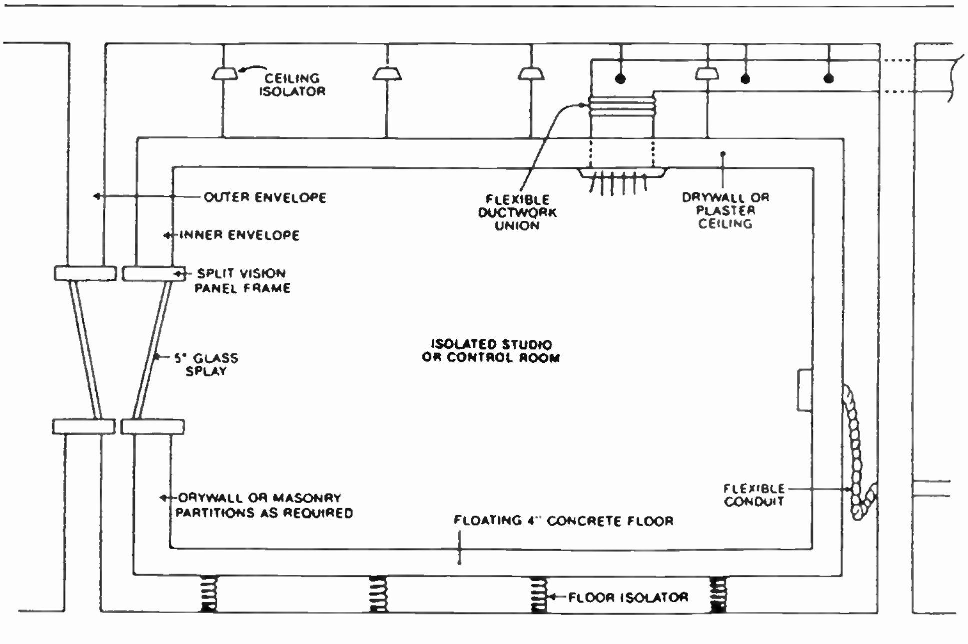


Figure 3.2-21. Floating construction technique.

Similar to the concept of STC, Impact Isolation Class (IIC) is a single number rating system to assess a barrier's effectiveness at arresting transmission of impact noise. The IIC method is based on the use of a standard *tapping machine* which supplies a known impact noise profile. SPL readings are then taken in one-third octave bands in the receiving room and compared with a standard contour to determine the IIC rating.

HVAC Considerations

Often the HVAC system in a room is the primary source for noise introduced into the room and also determines the level of acoustic leakage between rooms. With proper design, these limitations can in large measure be avoided.

A major contributor to HVAC noise is the fan noise itself which propagates down the supply ducts and subsequently enters the room. Ventilation ducts can be lined with absorbing material such as 1 in. glass fiber to reduce the noise at the end of the duct. Prefabricated duct silencers are also available which are placed in line with the duct and offer significant sound attenuation characteristics. Structure borne transmission of fan noise can be stopped by coupling the fan motor to the duct system via a canvas or rubber coupling to break up the vibration path.

The amount of noise in a duct system is also strongly dependent on the air velocity in the duct. Low noise design requires low air velocities which, for the same amount of total airflow, leads to the use of either multiple ducts or larger duct cross sections, both of which imply higher costs. Diffusor designs to minimize turbulence induced noise at the entry point to the room are also important.

The layout of supply and return ducts can be a hidden source of poor acoustic isolation between rooms. As shown in Figure 3.2-22, higher isolation is

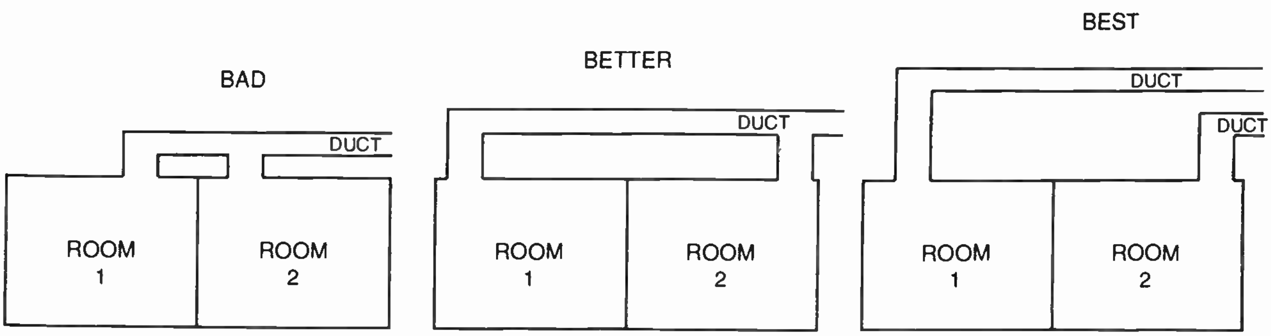
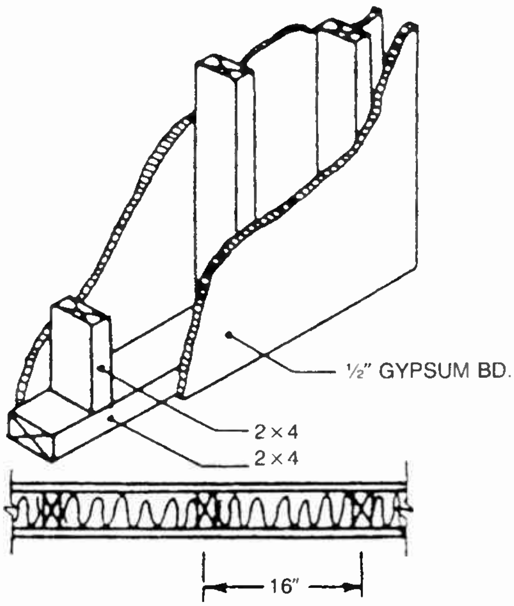


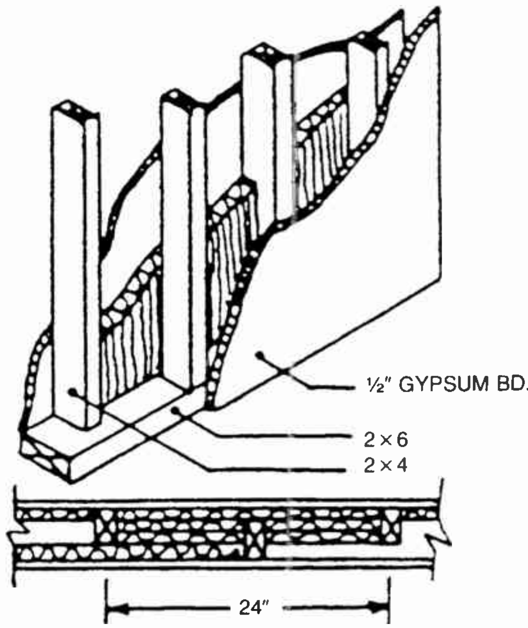
Figure 3.2-22. Effect of duct location on sound isolation.



Construction—standard stud partition

	Transmission Loss (dB)						STC
	125 Hz	250 Hz	500 Hz	1K Hz	2K Hz	4K Hz	
without glass fiber	15	27	36	42	47	40	35
with 3½" glass fiber	15	31	40	46	50	42	39

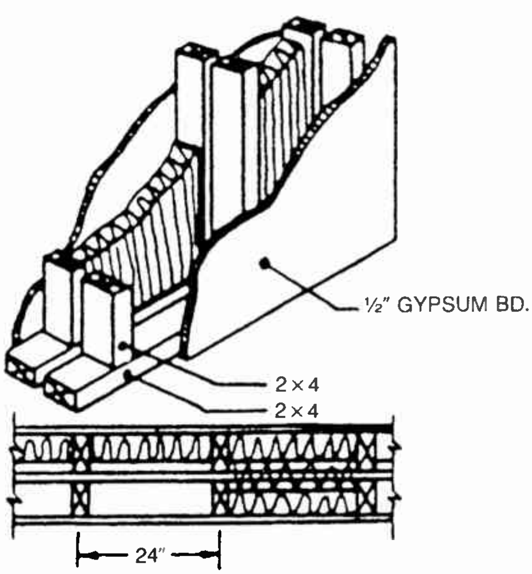
Figure 3.2-20(a). Standard stud partition. (Reprinted with the permission of Owens-Corning Fiberglass Corporation)



Construction—staggered stud partition

	Transmission Loss (dB)						STC
	125 Hz	250 Hz	500 Hz	1K Hz	2K Hz	4K Hz	
without glass fiber	22	23	36	46	52	41	38
with 3½" glass fiber	31	37	47	52	56	50	49

Figure 3.2-20(b). Staggered stud partition. (Reprinted with the permission of Owens-Corning Fiberglass Corporation)



Construction—double wall

	Transmission Loss (dB)						STC
	125 Hz	250 Hz	500 Hz	1K Hz	2K Hz	4K Hz	
without glass fiber	24	32	39	48	52	39	39
with 3½" glass fiber	29	43	54	63	66	52	52

Figure 3.2-20(c). Double wall partition. (Reprinted with the permission of Owens-Corning Fiberglass Corporation)

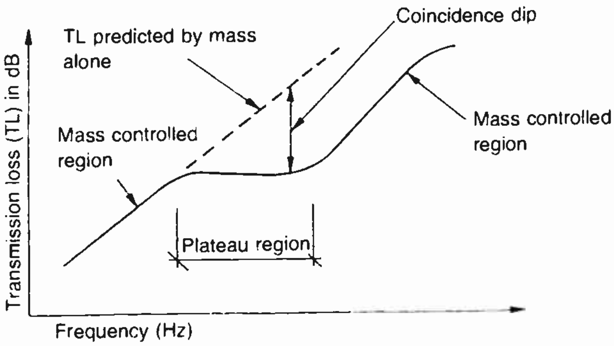


Figure 3.2-19. General transmission loss (TL) profile of solid materials.

(also known as a coincidence dip) and another mass controlled region above the plateau region. For rigid materials, especially at low frequencies, TL increases about 5 dB for each doubling of surface weight. In general, heavier materials have greater sound isolation capability.

Discontinuous Construction

Increasing the mass of a wall is an effective method of reducing the transmitted sound energy, but each doubling of mass will accomplish only about a 5 dB reduction in sound transmission. A more effective approach is the isolation of the two sides of the wall with an airspace and some form of resilient connection, often called discontinuous construction. It is important that the two sides of the wall not be rigidly tied together, as with wood studs, which will transmit sound energy as structure-borne sound.

Split or staggered wood studs (Figure 3.2-20C), lightweight metal studs, or resilient furring are all effective at isolating the two sides of a wall and reducing structure-borne sound. Within normal limits, the additional sound attenuation provided by an isolated airspace will also increase as the width of the airspace increases. Placing porous insulation (usually glass fiber) within the airspace can also increase the STC-rating by 3 to 6 dB, due to damping of the resonant coupling of the wall panels. The STC performance of common wall constructions is shown in Figure 3.2-20.

Windows

Achieving windows with high acoustic isolation requires a somewhat complex structure. A double pane with a 6 in. air gap between the panes is desirable. The glass must be resiliently mounted within the frame without cracks or air gaps. For best performance, the two panes should be different thicknesses and the perimeter of the air gap between the panes covered with sound absorbing material. Often one pane is angled with respect to the other, a characteristic which has more advantages in visual glare reduction than affecting the TL significantly.

Doors

High isolation doors are difficult to build and maintain since the gasketing and sealing of the door must be extremely precise and not degrade with time or wear and tear. Because of the difficulties of constructing and maintaining doors with high isolation, it is often more economical and reliable to create high isolation entry ways through use of a *sound lock,* that is, an outer door leading to a vestibule with an inner door.

The overall TL from the outside hall to the inside of the room via the vestibule can be very high even with only moderately stringent construction techniques. Also, by sequentially entering or exiting a sequence of doors, the room is never exposed to the total loss of isolation inherent to a single door partition.

Floating Construction

Floating construction is a combination of discontinuous construction and resilient mounting techniques. Floating floors are solid slab floors which are completely isolated from the structural floor by a resilient underlayment or resilient isolators. Walls may be built attached to the floating floor and a ceiling may be resiliently hung from the structural ceiling, resulting in an actual room-within-a-room with very high isolation possibilities. This is shown in Figure 3.2-21.

This type of construction can be extremely expensive but is sometimes the only avenue to achieving high levels of sound isolation, especially at low frequencies, in high noise environments or when very low ambient noise levels are required.

Vibration Isolation

Vibrations and sound energy produced by mechanical equipment can be transmitted throughout a building via vibration of the structure and reradiated as sound energy in a particular room. In general, vibrating equipment can be effectively isolated from the building structure by mounting the equipment on resilient mounts. The mass of the equipment and the compliance of the resilient mounting form a resonant system. Vibrations of the equipment at frequencies much higher than this resonant frequency will effectively be prevented from being transmitted into the building structure. As a rule of thumb, the resonant frequency should be one third or less than the lowest desired frequency of effective isolation. The lower the resonant frequency, the lower the level of transmitted vibrations for a given vibration frequency.

Impact Noise Reduction

Impact noise, as the name implies, refers to such mechanisms as footsteps, objects dropped on floors, slamming doors and so forth. Reduction of impact noise may in many cases be effectively treated by the obvious; a rug on the floor, castered chairs and other measures can drastically reduce impact problems by softening the blow. To obtain better improvements the same techniques to improve TL such as discontinuous construction can be effective.

Assessing Noise Levels

The Noise Criteria (NC) curves were developed through experimentation and knowledge of the sensitivities of the human hearing system to provide a single number figure of merit for maximum permissible noise level for a given activity. The NC number is given approximately by the value of the NC curve in the 1200–2400 Hz frequency band. Figure 3.2-18 shows the NC family of curves.

For assigning an NC rating to an arbitrary room:

- SPL readings are taken in the 8 octave bands from 63 to 8,000 Hz
- The NC rating is defined as the lowest value NC curve which lies wholly above the measured data.

For broadcast studios, a rating of NC-20 or less is desirable.

Airborne Noise Reduction

For sound energy striking a partition, one can define a transmission coefficient for the partition:

$$\tau = \frac{\textit{Energy transmitted through the partition}}{\textit{Energy incident on the partition}}$$

Transmission Loss (TL) is then defined as:

$$\text{TL} = 10 \log \left(\frac{1}{\tau}\right)$$

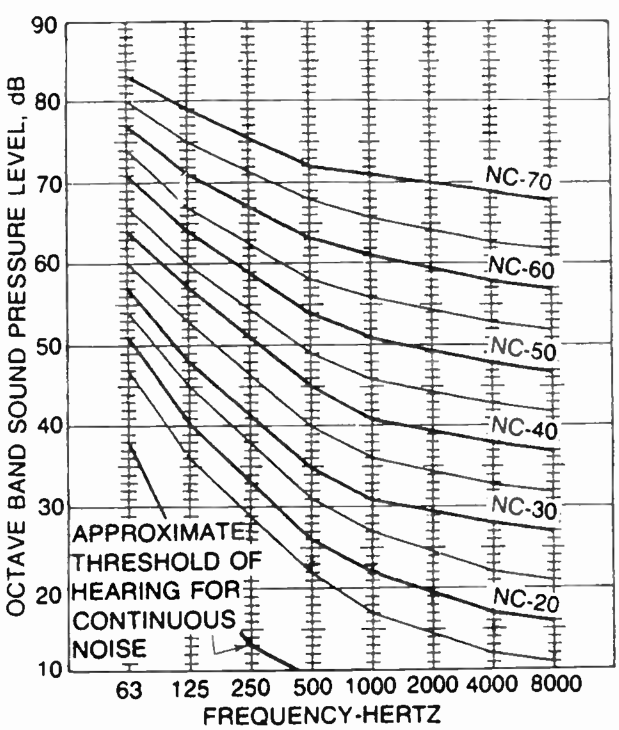


Figure 3.2-18. Noise Criteria (NC) curves. (L. Beranek, "Revised Criteria for Noise in Buildings." Noise Control, V. 3, N. 1, Jan. 1957, pp 19–27).[15]

and represents the decibel reduction of sound energy through the partition. The overall TL through a partition composed of several areas (for example, a wall with a door and window) can be calculated as follows:

$$\textit{Composite}\ \text{TL} = 10 \log \left[\frac{S_T}{\tau_1 s_1 + \tau_2 s_2 + \cdots \tau_n s_n}\right]$$

where

S_T = total surface area of partition
τ_n and s_n = the transmission coefficient and surface area of the nth element of the composite partition.

The disastrous effect of cracks and air gaps on achieving high TL partitions can easily be illustrated with this formula. For example, using $\tau = 1$ for a crack, consider the effect of a 1/8 in. crack under a door having a TL of 30 dB. The composite TL is then approximately 26 dB. Applying this same situation to a door with TL = 50 dB, yields a composite TL of approximately 28 dB, a 22 dB loss in isolation due to the air gap!

In partition design, the need for avoidance of cracks and gaps and the importance of gasketing and sealing cannot be over-stated—high sound isolation simply cannot be achieved if these factors are ignored or compromised.

Noise Reduction Between Rooms

Noise reduction (NR) is the actual difference in SPL measured in a room containing an offending noise source and the room under test. NR is determined by the area of the dividing partition (S), the total absorption in the receiving room (a), and the TL of the partition:

$$\text{NR} = \text{TL} + 10 \log \left(\frac{a}{S}\right)$$

Thus it can be seen that the actual isolation between rooms can vary both above and below the TL of the partition, although in practice, NR will typically be within 6 dB of the TL.

STC Rating

The Sound Transmission Class (STC) rating represents a single number figure of merit for overall acoustic isolation of a material. TL is plotted on ⅓ octave bands and compared with a standard contour curve to determine the STC rating. The higher the STC, the better the material for reducing sound transmission. While convenient as a figure of merit, in critical applications, such as broadcasting, it is always preferable to consider transmission loss at different frequencies to meet specific design objectives. There is no shortcut that can overcome the fact that transmission loss varies as a function of frequency.

Transmission Loss of Solid Materials

The TL of a solid wall tends to have three defined frequency regions as shown in Figure 3.2-19: a mass controlled region at low frequencies, a plateau region

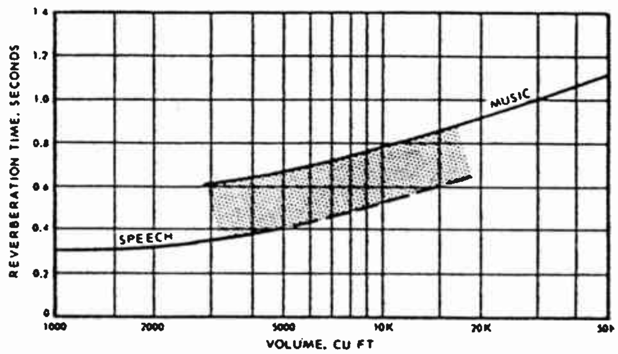


Figure 3.2-16. Optimum reverberation time (at mid-frequencies). (Reprinted with permission from book #1696 *Acoustic Techniques for Home and Studio*, 2nd Ed., by F. Alton Everest. Copyright 1984 by TAB Books, A Division of McGraw-Hill, Inc., Blue Ridge Summit, PA.

RT_{60} varies as a function of frequency. When not specified, mid frequencies around 500 Hz are usually the assumed frequency range. For critical listening applications, the optimum reverberation characteristic is sometimes deemed to be flat as a function of frequency. However, for a natural sounding environment with "warmth" it is often desirable to have higher RT_{60} at low frequencies compared to mid and high frequencies. Figure 3.2-17 shows the recommendations of the EBU in this regard.

Statistical Descriptors of Speech Intelligibility

Several statistical descriptors are used to classify the intelligibility of speech both in rooms and over audio networks. These include articulation index, speech transmission index, rapid speech transmission index and speech interference level.

Articulation index (AI) is the most often used of these metrics and can be measured using a simple sound source and sound level meter. It is the weighted proportion of speech which is usable to convey information and is calculated from the scores of a group of experienced listeners with normal hearing who write sentences, words or syllables read to them from specially-selected lists.

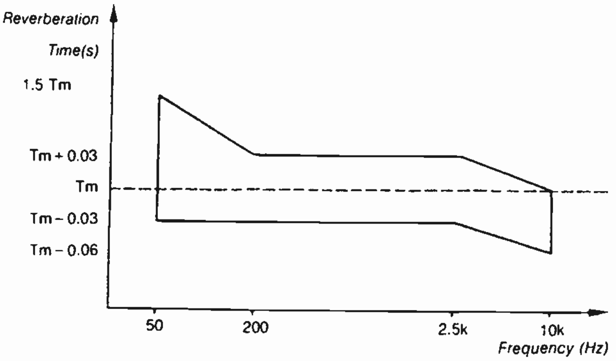


Figure 3.2-17. EBU recommendation on optimum reverberation time as a function of frequency. (Reprinted from EBU Technical Recommendation R22-1985: Acoustical properties of control rooms and listening rooms for the assessment of broadcast programs).

Speech transmission index (STI) is similar to AI but uses a modulation transfer function in its evaluation, and as such requires more sophisticated instrumentation, currently produced by only one manufacturer. It incorporates a procedure for expressing the loss in articulation produced by room reverberation.

The rapid speech transmission index (RASTI) is similar to STI, but uses fewer modulation frequencies and incorporates only speech and background sound levels in only two octave bands—500 Hz and 2000 Hz. Measurement of RASTI also requires sophisticated instrumentation.

A final descriptor, speech interference level (SIL) is the arithmetic average of the sound pressure level in four octave bands whose center frequencies are 500, 1000, 2000, and 4000 Hz.

NOISE CONTROL TECHNIQUES

Common sense is often the most important ingredient in acoustical room design and also the first forgotten concern when practical matters arise and introduce conflicts. In optimizing a room design for broadcast applications the following checklist identifies the highest priority items that should be considered, listed in approximate order of importance.

1. Select site for least noisy area
2. Perform noise survey to determine required acoustical isolation
3. Select position of room within building
4. Select sound isolation technique
5. Control noise within building including structure borne and airborne noise
6. Design room geometry for good diffusion and room mode spacing
7. Select and place sound absorbing material for optimum room acoustic response.

Numbers 1 through 5 refer to the control of noise and are intentionally a higher priority than criteria such as room shape and optimum reverberation time, especially for broadcast applications. A quiet room is both the most important design goal and also potentially the most expensive and difficult goal to achieve.

Below are four basic approaches to reducing noise within a room, listed in rough order of preference:

- Select room location
- Reduce output of noise source
- Design barrier partitions between noise source and room
- Reduce the noise energy existing within the room.

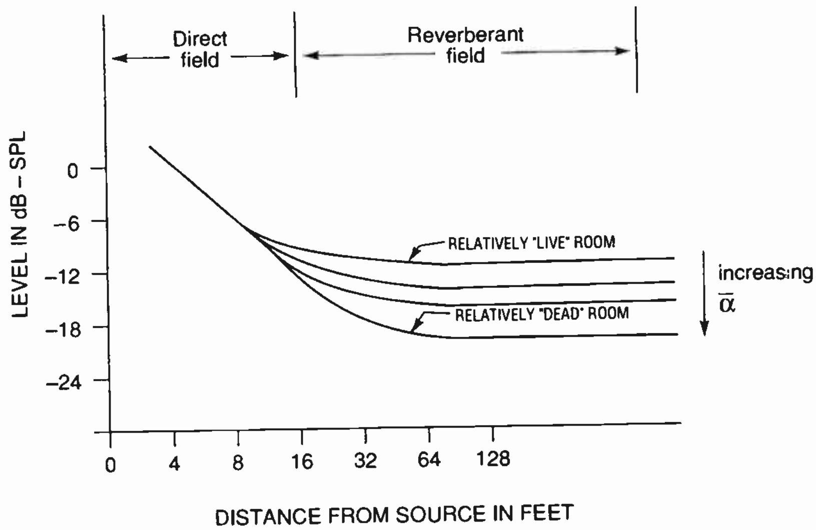


Figure 3.2-15. Sound attenuation as a function of distance and the amount of total room absorption.

Sound Decay in a Room

When a sound source stops in a room, the SPL at a given location will not decrease to zero instantaneously as in the free field case. Rather, the sound energy in the room will decay over a period of time due to reflected sound energy gradually dissipating as a result of the absorptive qualities of the room surfaces. This sound decay is called reverberation. The reverberation time of a room (RT_{60}) is defined to be the amount of time required for sound to decay 60 dB:

$$RT_{60} = \frac{0.049V}{S\overline{\alpha}} \quad (\text{when } \overline{\alpha} < 0.1)$$

where

V = room volume (ft^3)
S = surface area of room (ft^2)
$\overline{\alpha}$ = average absorption coefficient

This is the classic Sabine formula for RT_{60}.

For rooms in which the average absorption coefficient for all surfaces ($\overline{\alpha}$) is greater than 0.1 or where the absorption of various surfaces is significantly different, more complex equations govern. The reader is directed to *Acoustic Design and Noise Control*, by Michael Rettinger or other advanced texts on architectural acoustics.

In very large rooms or at high frequencies, the effect of excess absorption due to humidity and other effects must be taken into account:

$$RT_{60} = \frac{0.049V}{S\overline{\alpha} + 4\,m\mathrm{V}}$$

where m = excess sound attenuation in dB/ft (see Figure 3.2-2).

Limitations of RT_{60}

A basic premise of the RT_{60} formula is that the room exhibits a uniform rate of decay of sound. This in turn would require that the sound field is completely diffuse, an assumption generally more close to being true in large "live" rooms than small "dead" rooms. In small, very absorptive rooms, where all significant sound energy dies away in a few reflections, the validity of a statistically based tool like reverberation time becomes questionable. In these cases, the reflection profile itself must be considered on a more specific basis. In critical cases, in addition to addressing the statistical decay of the room versus frequency as represented by reverberation time, it is important to consider the strategic placement of absorptive, reflective and/or diffusive materials to provide control of first or second order reflections. For example, strategic reflection control may be important in a studio control room on the side walls and ceiling between the loudspeakers and listening position, as well as the front wall between the loudspeakers and the rear wall, which is often intentionally made diffusive according to several popular studio design philosophies.

Optimum reverberation time has received a lot of attention over the years. Being a subjective figure of merit, numbers that can be universally agreed upon for all circumstances will probably never exist. Certainly, the optimum RT_{60} varies with the size of the room and the intended application. For control rooms and listening rooms the European Broadcasting Union recommends an RT_{60} of .3 seconds ± .1 seconds at mid frequencies (200 Hz–2500 Hz). Figure 3.2-16 shows another example of optimum RT_{60} curves for small listening rooms and studios.

In general terms, most studio environments should have as little reverberation as possible. This is due to the simple fact that reverberation can be created electronically, but to date, it is virtually impossible to remove excess reverberation from an electronic signal.

the incident energy equally over a wide angular range and also corresponds to a widening of the energy received as a function of time.

Diffusing elements designed specifically for that purpose are available and the design of these units is quite refined. While traditional surface relief ornamentation, as referenced above, is useful as a diffusion element, it generally does not provide broad-bandwidth wide angle diffusion. An ideal diffusor would provide sound diffusion which is not a function of frequency, angle of incidence or observation angle. Examples of several commercially available products designed with these goals in mind are shown in Figure 3.2-14.

Based on the construction of a series of wells of different depths derived mathematically, the diffusors shown achieve a uniform angular distribution of reflected energy from a wide range of incident angles for mid and high frequencies.

Sound Level as a Function of Distance in an Enclosed Room

In an enclosed room, the situation is more complicated than the case of a free field—an environment with no reflected sound energy. The total SPL at any point in the room will be the result of contributions directly from the source (referred to as the direct sound field) and sound energy associated with multiple reflections from the room surfaces (referred to as the diffuse or reverberant sound field). The direct sound field varies inversely with the square of the distance from the source: the diffuse sound field is, by definition of being diffuse, equal at all points in the room. It can be theoretically shown that, for an enclosed room, at a distance r (in feet) from a sound source:

$$\mathrm{SPL} = \mathrm{PWL} + 10 \log \left[\frac{Q}{4\pi r^2} + \frac{4}{R} \right] + 10.5$$

where:

$R = \dfrac{S\overline{\alpha}}{1 - \overline{\alpha}}$

Q = Directivity of source

$\overline{\alpha}$ = Average absorption coefficient

S = Total room surface area

(R is sometimes called the Room Constant).

The graphical form of this equation is shown in Figure 3.2-15.

At distances remote from the sound source, the first fraction ($Q/4\pi r^2$) in the equation above becomes small as r^2 becomes large and the reverberant sound pressure level may be approximated as:

$$\mathrm{SPL} = \mathrm{PWL} - 10 \log R + 16.4$$

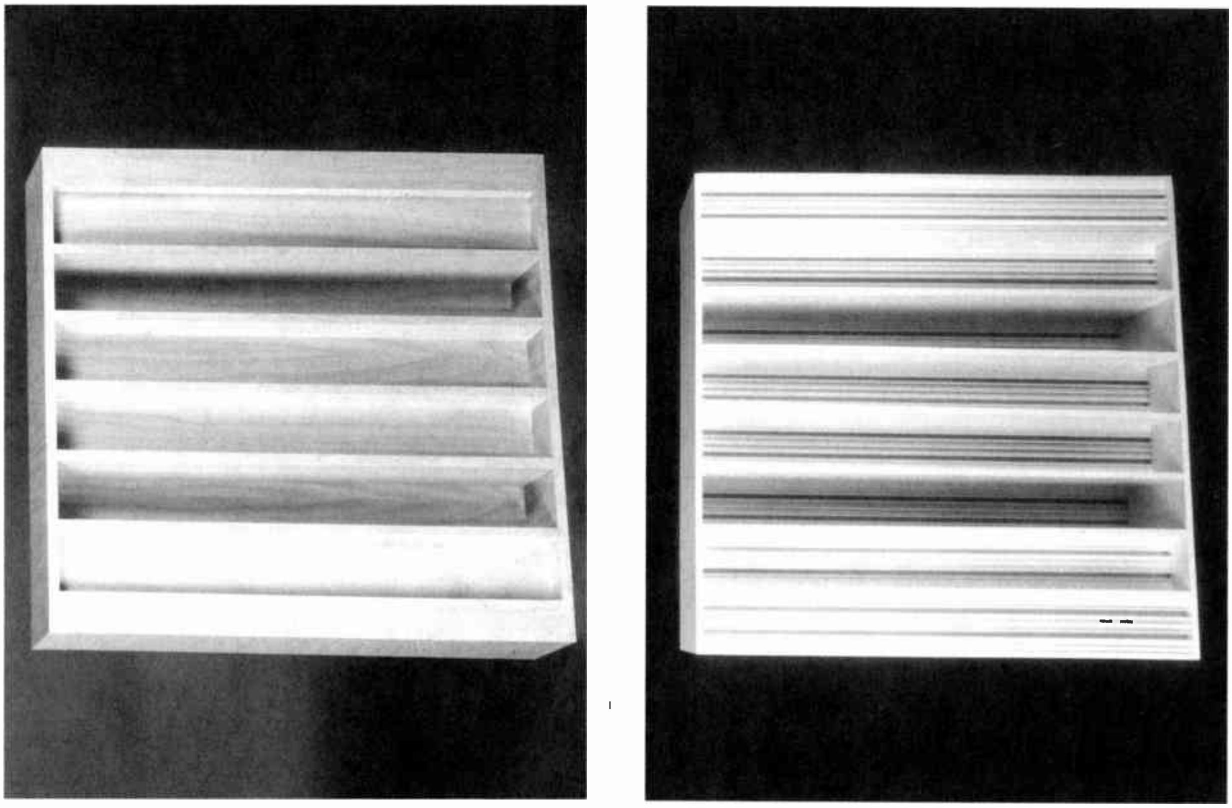


Figure 3.2-14. Examples of commercially available diffusors. (Photo courtesy of RPG Diffusor Systems.)

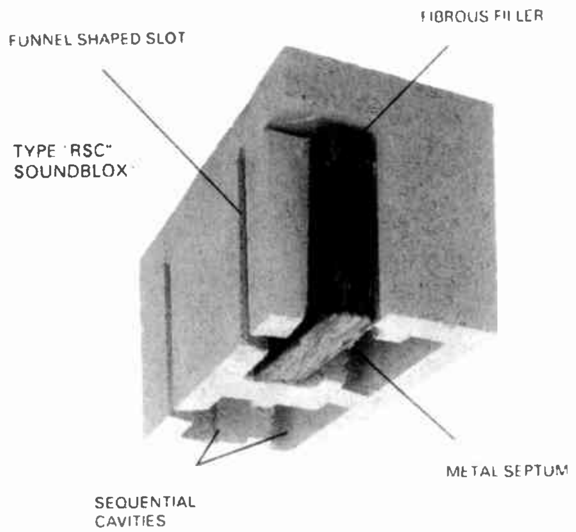


Figure 3.2-10. A cavity resonator used in construction. (Type RSC SOUNDBLOX® Courtesy of the Proudfoot Co. Inc.)

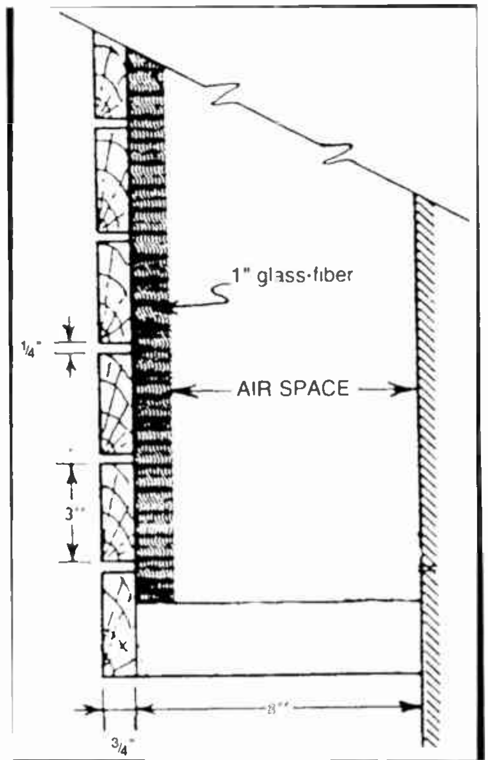


Figure 3.2-12. Example of a slit resonator.

success of many famous concert halls built in the 19th century can be largely attributed to the florid architectural features that offer diverse reflective and scattering properties for incident sound. Also, the intentionally irregular distribution of absorptive material in patches will increase diffusion. Increasing the randomness of the location of patches of absorptive material will also increase diffusion.

The range of possible effects of surface treatment on incident sound energy are shown in Figure 3.2-13. As shown in (b), a hard flat surface will reflect sound in a specular manner just as light is reflected from a mirror, following the relationship that the angle of incidence (relative to a line perpendicular to the surface) equals the angle of reflection and is equal in magnitude. At a location where both direct and reflected waves arrive, the reflected wave is a delayed replica of the incident wave. Similarly, sound absorptive material applied to a surface yields an attenuated and delayed replica of the incident wave as shown in (a). Finally, a diffusive surface, as shown in (c), reflects

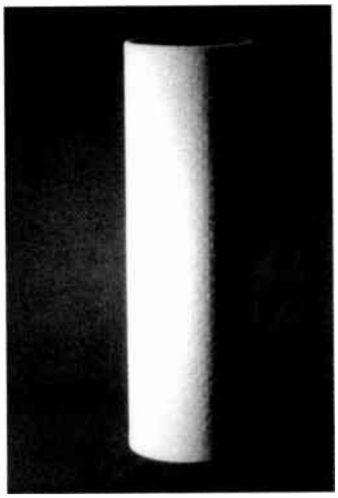


Figure 3.2-11. A portable cavity resonator. (TubeTrap® Photo courtesy of Acoustic Sciences Corp.)

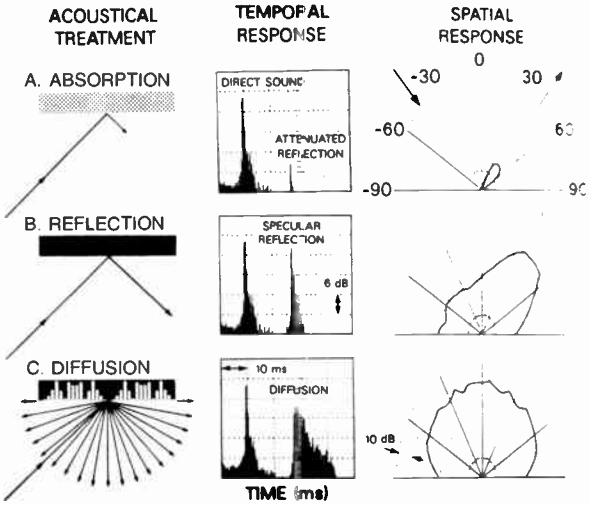
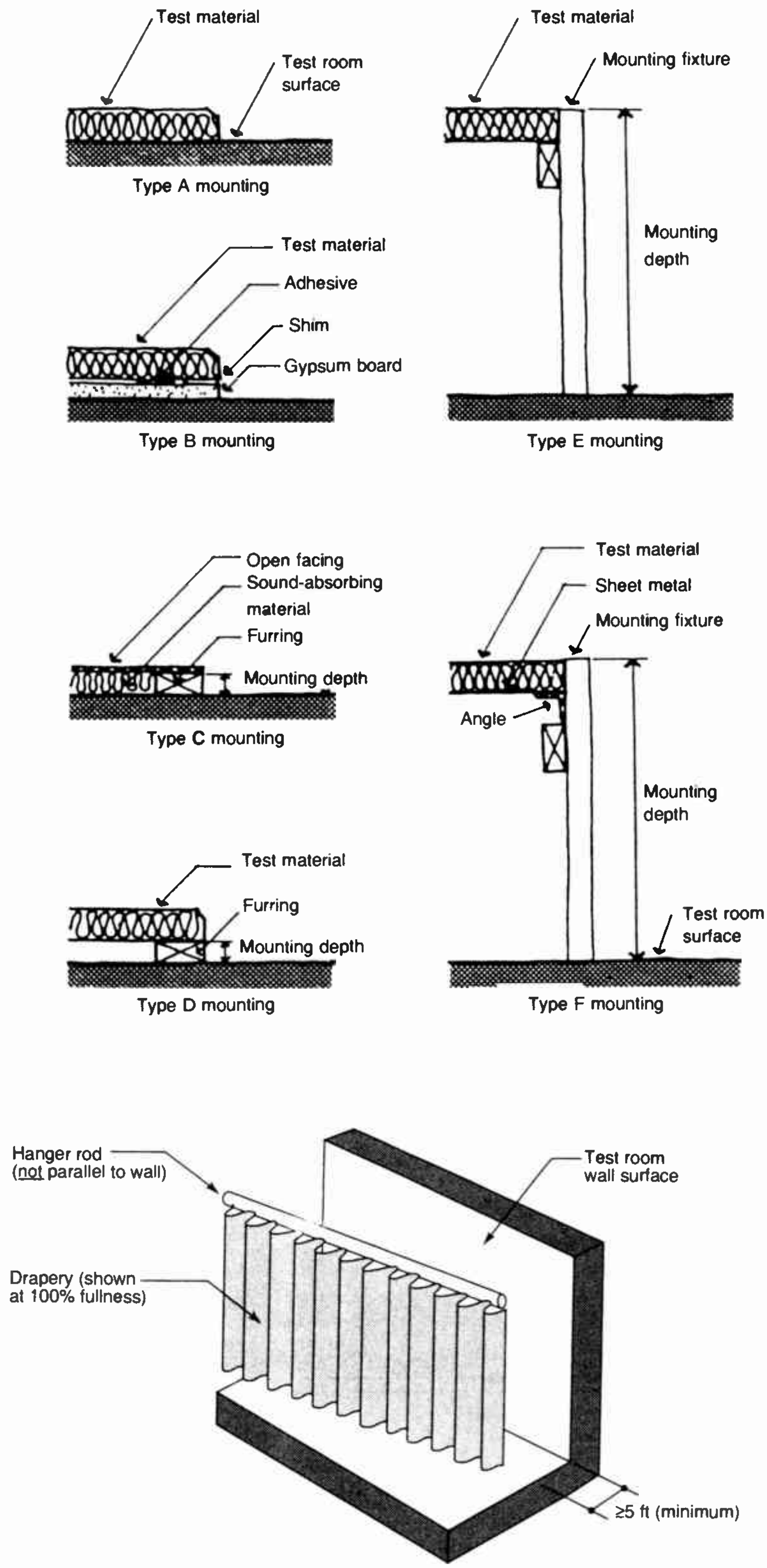


Figure 3.2-13. Reflection, absorption and diffusion from acoustic surface treatments. (Courtesy of RPG Diffusor Systems.)



Type H mounting
(to simulate space absorbers)

Figure 3.2-9. Standard mounting methods for measuring the absorption coefficient of a material. (ASTM C 423.)

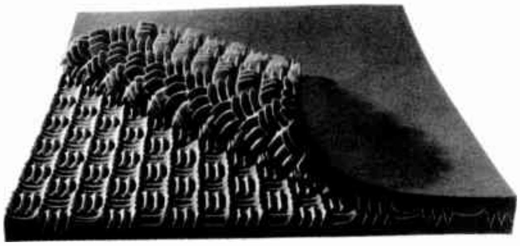


Figure 3.2-8. An open cell foam acoustical material. (SONEX photo courtesy of Illbruck, Inc.: SONEX Acoustical Products Div.)

Carpeting is often used as an absorber. However, absorption of carpeting at low frequencies is relatively poor. In general, a foam rubber or hair felt underlayment can improve low frequency absorption significantly. Typical absorption figures (with concrete as a reference) are listed in Table 3.2-3.

Effect of Mounting Method of Material on Absorption

Different methods of mounting a given material may give different absorption coefficient results. Tables of absorption coefficients must indicate the mounting method used when the absorption coefficient was measured to be useful.

In general terms, materials mounted against a hard backing will exhibit the lowest sound absorption coefficients. Sound absorption, particularly at low frequencies is increased by the presence of an airspace behind the absorptive material.

Figure 3.2-9 shows some of the various mounting designations used for standard measurements.

Acoustic Absorption Characteristics of People

Absorption by people and objects is specified in absorption tables as either the number of sabins per person or as an absorption coefficient provided by an audience based on normal seat and aisle spacings. Typical values are listed in Table 3.2-3.

Panel Absorbers

A panel with an enclosed air space behind it forms a resonant system with the air mass behaving as a spring and the panel as a mass. If the panel is thin, the resonant frequency and hence the frequency of maximum absorption coefficient can be shown to be:

$$f_{resonant} = \frac{170}{\sqrt{md}}$$

where

d = air space depth (in)
m = mass per unit area of panel (lb/sq. ft)

A broader absorption characteristic can be achieved by filling the air space with absorptive material, such as fiber glass.

Cavity Resonators

There are three basic types of cavity resonators:

- Individual units
- Perforated panels
- Slit resonators.

An example of an individual pre-fabricated cavity resonator is shown in Figure 3.2-10, which is essentially a slotted concrete block. Another type of commercially available individual unit is shown in Figure 3.2-11.

A perforated panel of significant thickness (greater than approximately 1/8 in.) spaced away from a rigid backing exhibits the absorptive behavior of a cavity resonator. The resonator has an approximate resonance frequency of:

$$f_{resonant} = 200\sqrt{\frac{p}{dt}}$$

where

p = percentage of open area of panel
d = air space depth (in.)
t = panel thickness + 0.8 × hole diameter (in.)

A slit resonator consists of a number of slats spaced away from a rigid backing with air spaces (slots) in between the slats as shown in Figure 3.2-12.

The resonant frequency can be calculated as follows:

$$f_{resonant} = 2{,}160\sqrt{\frac{s}{dD(w + s)}}$$

where

s = Width of slot (in.)
d = Thickness of slat (in.)
D = Depth of air space (in.)
w = Width of slat (in.)

In the example in Figure 3.2-12, a resonant frequency of about 250 Hz results. Broader absorption characteristics can be achieved by using slats and slots of varying widths or non-parallel air spaces.

Diffusion

Diffusion of sound in a room refers to the extent to which sound energy is uniformly distributed throughout a room. Making an analogy with baking a cake, diffusion would represent an indicated measure of how well the batter was mixed and individual ingredients of the recipe were dispersed throughout the mixture. In pursuing good room acoustics, maximum diffusion is desirable, all other things being equal. In a purely diffuse sound field:

- At a given location, sound waves are equally likely to be traveling in any direction
- The sound pressure will be equal at all locations throughout the room.

These criteria are determined by the pattern of reflections within the room. Diffusion in a room can be maximized in several ways. The introduction of oddly shaped protrusions aids in increasing diffusion. The

Table 3.2-3 (continued)

Material	Sound Absorption Coefficient						NRC*
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	
Ceilings[6, 8–10]† (continued)							
Sound-Absorbing:							
42. Acoustical board, ¾-in thick, in suspension system (mtg. E)	0.76	0.93	0.83	0.99	0.99	0.94	0.95
43. Shredded-wood fiberboard, 2 in thick on lay-in grid (mtg. E)	0.59	0.51	0.53	0.73	0.88	0.74	0.65
44. Thin, porous sound-absorbing material, ¾-in-thick (mtg. B)	0.10	0.60	0.80	0.82	0.78	0.60	0.75
45. Thick, porous sound-absorbing material, 2 in thick (mtg. B), or thin material with airspace behind (mtg. D)	0.38	0.60	0.78	0.80	0.78	0.70	0.75
46. Sprayed cellulose fibers, 1 in thick on concrete (mtg. A)	0.08	0.29	0.75	0.98	0.93	0.76	0.75
47. Glass-fiber roof fabric, 12 oz/yd²	0.65	0.71	0.82	0.86	0.76	0.62	0.80
48. Glass-fiber roof fabric, 37½ oz/yd² (*Note:* Sound-reflecting at most frequencies.)	0.38	0.23	0.17	0.15	0.09	0.06	0.15
49. Polyurethane foam, 1 in thick, open cell, reticulated	0.07	0.11	0.20	0.32	0.60	0.85	0.30
50. Parallel glass-fiberboard panels, 1 in thick by 18 in deep, spaced 18 in apart, suspended 12 in below ceiling	0.07	0.20	0.40	0.52	0.60	0.67	0.45
51. Parallel glass-fiberboard panels, 1 in thick by 18 in deep, spaced 6½-in apart, suspended 12 in below ceiling	0.10	0.29	0.62	1.12	1.33	1.38	0.85
Seats and Audience[1, 5, 7, 9]‡							
52. Fabric well-upholstered seats, with perforated seat pans, unoccupied	0.19	0.37	0.56	0.67	0.61	0.59	
53. Leather-covered upholstered seats, unoccupied§	0.44	0.54	0.60	0.62	0.58	0.50	
54. Audience, seated in upholstered seats§	0.39	0.57	0.80	0.94	0.92	0.87	
55. Congregation, seated in wooden pews	0.57	0.61	0.75	0.86	0.91	0.86	
56. Chair, metal or wood seat, unoccupied	0.15	0.19	0.22	0.39	0.38	0.30	
57. Students, informally dressed, seated in tablet-arm chairs	0.30	0.41	0.49	0.84	0.87	0.84	
58. Person, adult (total number of sabins)[13]	2.5	3.5	4.2	4.6	5.0	5.0	
Miscellaneous[3, 9, 11]							
59. Gravel, loose and moist, 4 in thick	0.25	0.60	0.65	0.70	0.75	0.80	0.70
60. Grass, marion bluegrass, 2 in high	0.11	0.26	0.60	0.69	0.92	0.99	0.60
61. Snow, freshly fallen, 4 in high	0.45	0.75	0.90	0.95	0.95	0.95	0.90
62. Soil, rough	0.15	0.25	0.40	0.55	0.60	0.60	0.45
63. Trees, balsam firs, 20 ft² ground area per tree, 8 ft high	0.03	0.06	0.11	0.17	0.27	0.31	0.15
64. Water surface (swimming pool)	0.01	0.01	0.01	0.02	0.02	0.03	0.00

* NRC (noise reduction coefficient) is a single-number rating of the sound absorption coefficients of a material. It is an average that only includes the coefficients in the 250 to 2000 Hz frequency range and therefore should be used with caution.

†Refer to manufacturer's catalogs for absorption data which should be from up-to-date tests by independent acoustical laboratories according to current ASTM procedures.

‡Coefficients are per square foot of seating floor area or per unit. Where the audience is randomly spaced (e.g., courtroom, cafeteria), mid-frequency absorption can be estimated at about 5 sabins per person. To be precise, coefficients per person must be stated in relation to spacing pattern.

§The floor area occupid by the audience must be calculated to include an *edge effect* at aisles. For an aisle bounded on both sides by audience, include a strip 3 ft wide; for an aisle bounded on only one side by audience, including a strip 1½ ft wide. No edge effect is used when the seating abuts walls or balcony fronts (because the edge is shielded). The coefficients are also valid for orchestra and choral areas at 5 to 8 ft² per person. Orchestra areas include people, instruments, music racks, etc. No edge effects are used around musicians.

Test Reference

"Standard Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method," ASTM C 423. Available from American Society for Testing and Materials (ASTM), 1916 Race Street, Philadelphia, PA 19103.

Sources

1. L. L. Beranek, "Audience and Chair Absorption in Large Halls," *Journal of the Acoustical Society of America*, January 1969.
2. A. N. Burd et al., "Data for the Acoustic Design of Studios," British Broadcasting Corporation, BBC Engineering Monograph no. 64, November 1966.
3. E. J. Evans and E. N. Bazley, "Sound Absorbing Materials," H. M. Stationery Office, London, 1964.
4. R. A. Hedeen, *Compendium of Materials for Noise Control*, National Institute for Occupational Safety and Health (NIOSH), Publication no. 80-116, Cincinnati, Ohio, May 1980. (Contains sound absorption data on hundreds of commercially available materials.)
5. H. F. Kingsbury and W. J. Wallace, "Acoustic Absorption Characteristics of People," *Sound and Vibration*, December 1968.
6. T. Mariner, "Control of Noise by Sound-Absorbent Materials," *Noise Control*, July 1957.
7. J. E. Moore and R. West, "In Search of an Instant Audience," *Journal of the Acoustical Society of America*, December 1970.
8. R. Moulder and J. Merrill, "Acoustical Properties of Glass Fiber Roof Fabrics," *Sound and Vibration*, October 1983.
9. "Performance Data, Architectural Acoustical Materials," Acoustical and Insulating Materials Association (AIMA). (This bulletin was published annually from 1941 to 1974.)
10. W. E. Purcell, "Materials for Noise and Vibration Control," *Sound and Vibration*, July 1982.
11. W. Siekman, "Outdoor Acoustical Treatment: Grass and Trees," *Journal of the Acoustical Society of America*, October 1969.
12. "Sound Conditioning with Carpet," The Carpet and Rug Institute, Dalton, Ga., 1970.
13. Knudsen, V. & Harris, C., "Acoustical Designing in Architecture," American Institute of Physics, 1978.

Table 3.2-3
Absorption coefficients of various materials. (Reprinted with permission from *Architectural Acoustics*, by M. D. Egan. Copyright 1988 by McGraw-Hill).

Material	Sound Absorption Coefficient						
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	NRC*
Walls(1–3, 9, 12)							
Sound-Reflecting:							
1. Brick, unglazed	0.02	0.02	0.03	0.04	0.05	0.07	0.05
2. Brick, unglazed and painted	0.01	0.01	0.02	0.02	0.02	0.03	0.00
3. Concrete, rough	0.01	0.02	0.04	0.06	0.08	0.10	0.05
4. Concrete block, painted	0.10	0.05	0.06	0.07	0.09	0.08	0.05
5. Glass, heavy (large panes)	0.18	0.06	0.04	0.03	0.02	0.02	0.05
6. Glass, ordinary window	0.35	0.25	0.18	0.12	0.07	0.04	0.15
7. Gypsum board, ½-in-thick (nailed to 2 × 4s, 16 in oc)	0.29	0.10	0.05	0.04	0.07	0.09	0.05
8. Gypsum board, 1 layer, ⅝-in-thick (screwed to 1 × 3s, 16 in oc with airspaces filled with fibrous insulation)	0.55	0.14	0.08	0.04	0.12	0.11	0.10
9. Construction no. 8 with 2 layers of ⅝-in-thick gypsum board	0.28	0.12	0.10	0.07	0.13	0.09	0.10
10. Marble or glazed tile	0.01	0.01	0.01	0.01	0.02	0.02	0.00
11. Plaster on brick	0.01	0.02	0.02	0.03	0.04	0.05	0.05
12. Plaster on concrete block (or 1 in thick on lath)	0.12	0.09	0.07	0.05	0.05	0.04	0.05
13. Plaster on lath	0.14	0.10	0.06	0.05	0.04	0.03	0.05
14. Plywood, ⅜-in paneling	0.28	0.22	0.17	0.09	0.10	0.11	0.15
15. Steel	0.05	0.10	0.10	0.10	0.07	0.02	0.10
16. Venetian blinds, metal	0.06	0.05	0.07	0.15	0.13	0.17	0.10
17. Wood, ¼-in paneling, with airspace behind	0.42	0.21	0.10	0.08	0.06	0.06	0.10
18. Wood, 1-in paneling with airspace behind	0.19	0.14	0.09	0.06	0.06	0.05	0.10
Sound-Absorbing:							
19. Concrete block, coarse	0.36	0.44	0.31	0.29	0.39	0.25	0.35
20. Lightweight drapery, 10 oz/yd², flat on wall (*Note:* Sound-reflecting at most frequencies.)	0.03	0.04	0.11	0.17	0.24	0.35	0.15
21. Mediumweight drapery, 14 oz/yd², draped to half area (i.e., 2 ft of drapery to 1 ft of wall)	0.07	0.31	0.49	0.75	0.70	0.60	0.55
22. Heavyweight drapery, 18 oz/yd², draped to half area	0.14	0.35	0.55	0.72	0.70	0.65	0.60
23. Fiberglass fabric curtain, 8½ oz/yd², draped to half area (*Note:* The deeper the airspace behind the drapery (up to 12 in), the greater the low-frequency absorption.)	0.09	0.32	0.68	0.83	0.39	0.76	0.55
24. Shredded-wood fiberboard, 2 in thick on concrete (mtg. A)	0.15	0.26	0.62	0.94	0.64	0.92	0.60
25. Thick, fibrous material behind open facing	0.60	0.75	0.82	0.80	0.60	0.38	0.75
26. Carpet, heavy, on ⅝-in perforated mineral fiberboard with airspace behind	0.37	0.41	0.63	0.85	0.96	0.92	0.70
27. Wood, ½-in paneling, perforated 3⁄16-in-diameter holes, 11% open area, with 2½-in glass fiber in airspace behind	0.40	0.90	0.80	0.50	0.40	0.30	0.65
Floors(9, 11)							
Sound-Reflecting:							
28. Concrete or terrazzo	0.01	0.01	0.02	0.02	0.02	0.02	0.00
29. Linoleum, rubber, or asphalt tile on concrete	0.02	0.03	0.03	0.03	0.03	0.02	0.05
30. Marble or glazed tile	0.01	0.01	0.01	0.01	0.02	0.02	0.00
31. Wood	0.15	0.11	0.10	0.07	0.06	0.07	0.10
32. Wood parquet on concrete	0.04	0.04	0.07	0.06	0.06	0.07	0.05
Sound Absorbing:							
33. Carpet, heavy, on concrete	0.02	0.06	0.14	0.37	0.60	0.65	0.30
34. Carpet, heavy, on foam rubber	0.08	0.24	0.57	0.69	0.71	0.73	0.55
35. Carpet, heavy, with impermeable latex backing on foam rubber	0.08	0.27	0.39	0.34	0.48	0.63	0.35
36. Indoor-outdoor carpet	0.01	0.05	0.10	0.20	0.45	0.65	0.20
Ceilings(6, 8–10)†							
Sound-Reflecting:							
37. Concrete	0.01	0.01	0.02	0.02	0.02	0.02	0.00
38. Gypsum board, ½-in-thick	0.29	0.10	0.05	0.04	0.07	0.09	0.05
39. Gypsum board, ½-in-thick, in suspension system	0.15	0.10	0.05	0.04	0.07	0.09	0.05
40. Plaster on lath	0.14	0.10	0.06	0.05	0.04	0.03	0.05
41. Plywood, ⅜-in thick	0.28	0.22	0.17	0.09	0.10	0.11	0.15

(continued)

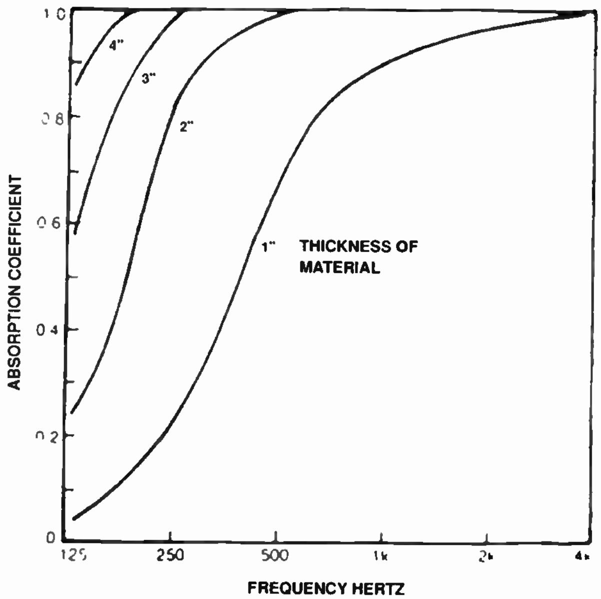


Figure 3.2-5. Effect of thickness on absorption characteristics of glass fiber (3 lb/cu ft density) mounted directly on a hard surface. (Reprinted, with permission, from book #3096 *The Master Handbook of Acoustics*, 2nd Edition, by F. Alton Everest. Copyright 1981, 1989 by TAB Books, A Division of McGraw-Hill Inc., Blue Ridge Summit, PA, 17294-0850. (1-800-233-1128.)

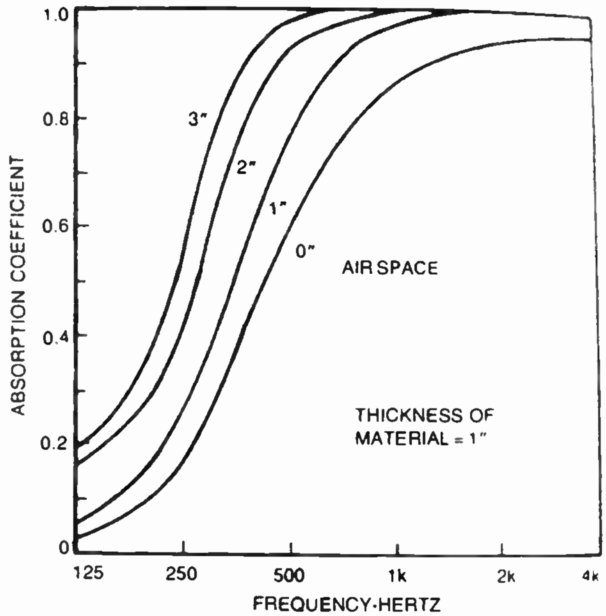


Figure 3.2-6. Effect of airspace on absorption characteristics of glass fiber (3 lb/cu ft density). Reprinted, with permission, from book #3096 *The Master Handbook of Acoustics*, 2nd Edition, by F. Alton Everest. Copyright 1981, 1989 by TAB Books, A Division of McGraw-Hill Inc., Blue Ridge Summit, PA. 17294-0850. (1-800-233-1128.)

as a function of an airspace between the glass fiber and the wall. Figure 3.2-7 shows that the effect of packing density on the absorption performance is rather small.

Drapes and curtains are also often used to attenuate reflected sound energy. However, drapes also tend to have high absorption at high frequencies and low absorption at low frequencies. Listed below are some guidelines to increase the low frequency absorption:

- Use heavy base material with a lining
- Use 100% to 200% gathering of the drape
- Hang at least 6 in. from the wall.

Examples of the absorption of drapes are listed in Table 3.2-3.

An example of a commercially available cellular foam type of absorber is shown in Figure 3.2-8.

Although irregular configurated foam is popular for its appearance, the user should be aware that the geometric pattern of wedges and cavities does not increase the sound absorption of cellular foam except at very high frequencies, where the depth of the wedges is equal to $\lambda/2$, typically above 10,000 Hz. Foam of uniform thickness equal to the average thickness will be equally effective at most frequencies. Installing cellular foam over an airspace will be far more effective at increasing sound absorption coefficients than varying the surface of the foam.

Only open-cell (soft) foams are acoustically absorptive. Rigid, closed-cell foam, although a good thermal insulator, does not provide significant acoustical absorption.

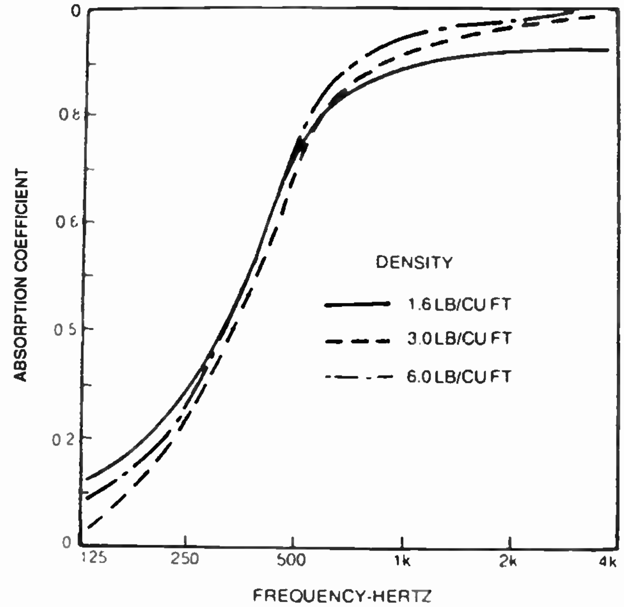


Figure 3.2-7. Effect of packing density on absorption characteristics of glass fiber (1 in. thickness). Reprinted, with permission, from book #3096 *The Master Handbook of Acoustics*, 2nd Edition, by F. Alton Everest. Copyright 1981, 1989 by TAB Books, A Division of McGraw-Hill, Inc., Blue Ridge Summit, PA.

sign, an open window is considered a perfect absorber since no energy that strikes the opening is reflected back into the room.

The total absorption of a surface is defined as the product of its absorption coefficient and the surface area. The unit of absorption is the sabin, named for the early 20th Century acoustician, Wallace Clement Sabine.

In many applications, it is convenient to define the average absorption coefficient of the room, $\bar{\alpha}$. $\bar{\alpha}$ is defined as the total absorption in the room (in sabins) divided by the total surface area:

$$\bar{\alpha} = \frac{\alpha_1 s_1 + \alpha_2 s_2 + \cdots \alpha_n s_n}{s_1 + s_2 + \cdots s_n} = \frac{\sum_{i=1}^{n} \alpha_i s_i}{s_T}$$

where α_n and s_n are the absorption coefficient and surface area of a portion of the room, and s_T is the total surface area of the room.

The average absorption coefficient is accurate only for the frequencies at which the absorption coefficient values are valid.

Absorption coefficients vary as a function of frequency and are usually listed in tables at 125, 250, 500, 1,000 2,000, and 4,000 Hz.

Often materials are specified by their Noise Reduction Coefficient (NRC) value. NRC was intended to be a convenient single number index of average sound absorbing efficiency. It is defined as the arithmetic average of absorption coefficients at 250, 500, 1,000 and 2,000 Hz, rounded off to the closest multiple of .05.

$$\text{NRC} = \frac{\alpha_{250} + \alpha_{500} + \alpha_{1000} + \alpha_{2000}}{4}$$

While this has some merit, NRC has the considerable disadvantage that low (and high) frequency absorption is not considered, an important criteria to consider in critical broadcasting applications. Also, the variation of α with frequency is an important consideration as well in critical applications where optimizing the total absorption in a room at all frequencies is important. Since NRC is an unweighted average, two materials with the same NRC could have drastically different absorption versus frequency characteristics. Wherever possible, absorption in a room should be analyzed separately in the various frequency ranges of interest.

Table 3.2-3 shows the absorption coefficients of various commercial sound absorbers and general building materials.

Sound Absorbing Material Applications

Echo Control

Sound absorbing material can be used very effectively in specific trouble areas to stop echoes or flutter echoes (delayed sound reflections of sufficient intensity to be heard discretely above the general reverberant sound level).

Noise Reduction

Sound absorbing material can be used to control noise within a room by lessening the amount of reverberant (reflected) energy present.

$$\textit{Noise reduction (in dB)} = 10 \log \left(\frac{a_{after}}{a_{before}} \right)$$

where a is the total absorption in the room in sabins before and after room treatment.

Note that the total absorption must be doubled to lower the noise 3 dB and doubled again to achieve a total 6 dB reduction. A practical limit is quickly reached in attempting to achieve more than 6 to 10 dB of noise reduction using this technique.

Reverberation Control

The "liveness" or "deadness" of a room can be controlled by the introduction of sound absorbing material. This is perhaps the most common use of sound absorbing material.

Types of Absorbers

Porous absorbers. Characterized by a material with deep pores and cavities. Sound energy entering the pores is dissipated by frictional and viscous resistance and/or vibrations of fibers of the material. Glass fiber is a typical porous absorber.

Panel absorbers. Sound energy forces a panel into vibration. The vibrational activity converts the sound energy into heat.

Cavity (Helmholtz) resonator. Analogous to blowing air across the mouth of a jug. At the resonant frequency, air in the jug neck vibrates back and forth as a single air mass. Sound energy is dissipated by frictional resistance in and around the neck.

Examples of Porous Absorbers

Acoustical blankets such as glass fibers are common porous absorbers. The absorption of a given material depends on its thickness, density and relative porosity. In general,

- Increasing thickness increases absorption (mainly at low frequencies)
- Blankets of relatively low density have increased absorption as the density increases
- Blanket must be composed of interconnected open pores (closed cell foam is a poor absorber). As a guideline, if a blanket will pass smoke under moderate pressure, it will probably be a good absorber.

Glass fiber is certainly one of the workhorse materials for sound absorption applications. It is interesting to note the effect of thickness and mounting arrangement on the absorptive qualities of this material. Figure 3.2-5 shows that increased thickness mainly affects the low frequency performance. (High frequency absorption is more a function of the surface texture and relatively independent of thickness). Figure 3.2-6 shows the effect of increased low frequency absorption

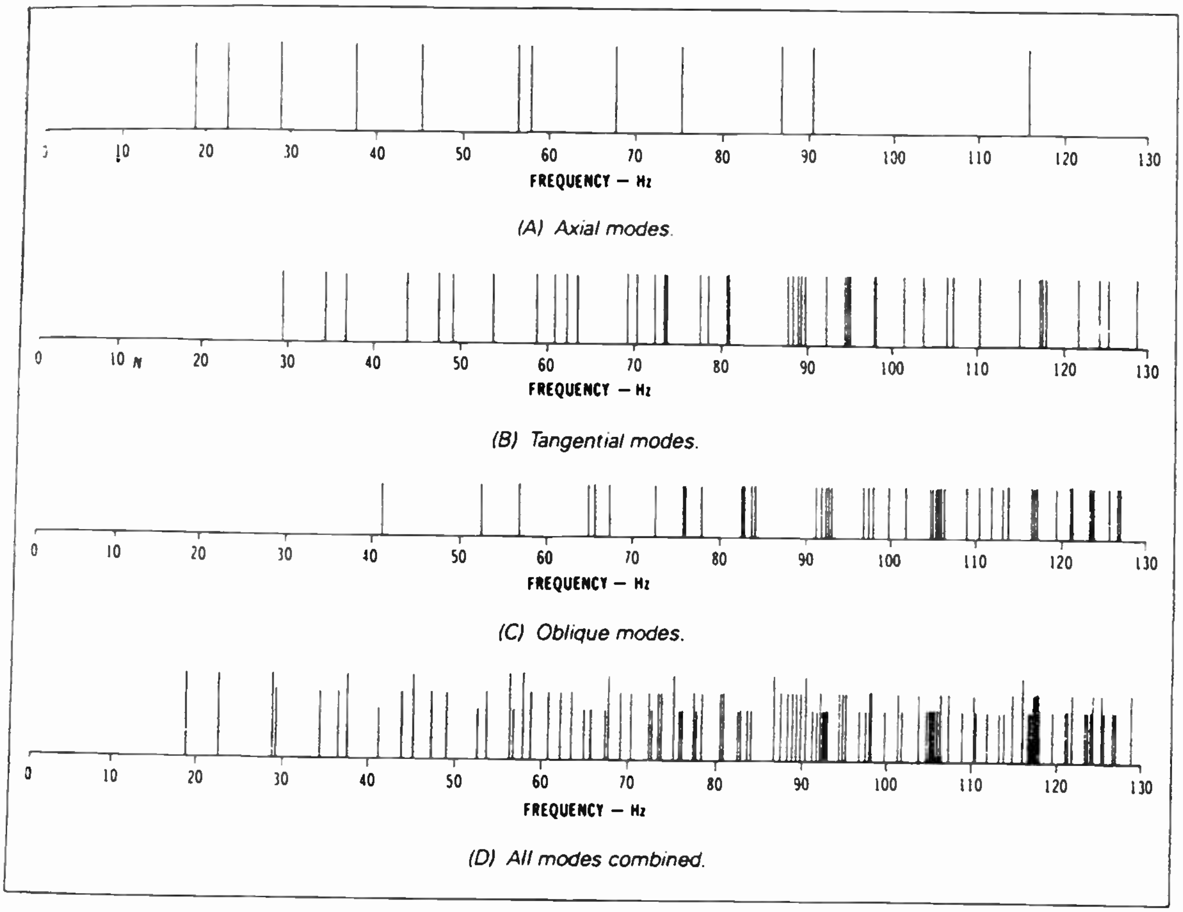


Figure 3.2-4. Modal frequencies for a room 30 ft × 25 ft × 19.5 ft (Glen Ballou: *Handbook for Sound Engineers, the New Audio Cyclopedia*, H. Sams, 1987).

pattern will lead to an overall optimized modal pattern since the tangential and oblique modes will tend to fill in the gaps between the axial modes.

Because of the resonance problems of small rooms, minimum sizes have been investigated over the years. The European Broadcast Union recommends a preferred volume of 80 cubic meters (2825 ft^3) for control rooms and 100 cubic meters (3500 ft^3) for listening rooms. Less than 40 cubic meters (1410 ft^3) is not recommended for use due to mode spacing problems.

Use of Nonparallel Walls

When the room is not rectangular, room mode problems do not vanish, they just become difficult to calculate. The modal distribution is certainly affected by room shape changes, but room resonances are not eliminated since the basic presence of room modes is more associated with volume as opposed to room shape. Nonparallel walls have been successfully used to promote good diffusion at higher frequencies and prevent flutter echoes. Flutter echoes are produced by repeated reflections from two flat parallel surfaces sufficiently distant from each other to produce a distinct echo. Usually an angle of 5° or a splay of one foot in 12 running feet is enough to destroy flutter echo problems. However, other more cost effective techniques exist to solve these problems.

Absorption of Sound

In general, when sound waves strike a surface:

- A portion of the energy is transmitted through the surface
- A portion of the energy is reflected back into the room
- A portion of the energy is absorbed.

The figure of merit for the absorptive quality of a material is called the absorption coefficient, α.

$$\alpha = \frac{\textit{Energy absorbed by the surface}}{\textit{Energy incident on the surface}}$$

α ranges between 0 and 1, 0 being a perfect reflector and 1 being a perfect absorber. In practical terms, absorption coefficients of surfaces are used in room design to aid in controlling reflections. For room de-

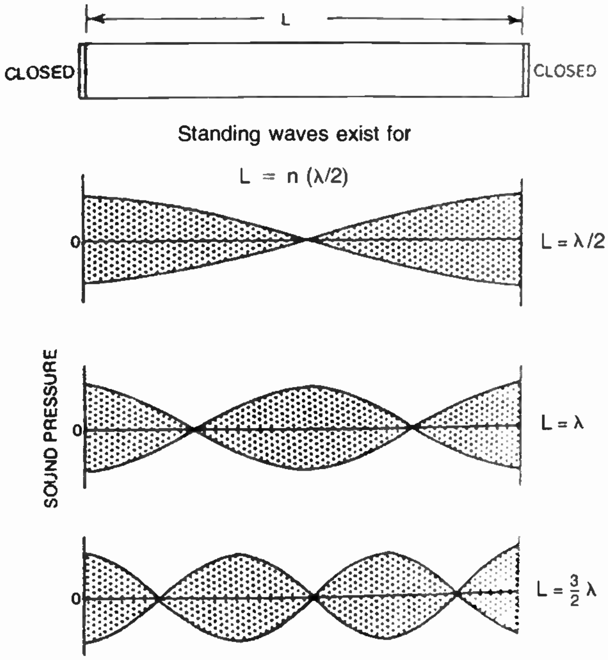


Figure 3.2-3. Standing waves in a closed tube.

three-dimensional instead of one-dimensional. Three types of resonance or room modes can be described:

- **Axial modes.** Resonant condition involving two parallel surfaces of room.
- **Tangential modes.** Resonant condition involving four surfaces and parallel to two surfaces of room.
- **Oblique modes.** Resonant condition involving all six surfaces of room.

The axial, tangential and oblique room modes for a rectangular room can be calculated as follows:

$$F = \frac{C}{2}\left[\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2\right]^{1/2}$$

where:

L, W, H = length, width and height of room
C = the speed of sound
p, q, r = integers 0, 1, 2, 3, . . .

A set of p, q and r represents one mode of vibration.

For the axial modes, only one of p, q or r will be non-zero. For tangential modes, two of p, q and r will be non-zero. Oblique modes have p, q and r all non-zero.

The above formula can be simplified for the axial modes:

$$\textit{Axial mode frequencies} = \frac{Cp}{2L}, \frac{Cq}{2W} \text{ and } \frac{Cr}{2H}$$

$$\text{for } p, q, r = 1, 2, 3, \ldots$$

At lower frequencies, modal frequencies are spaced farther apart. As frequency is increased, the modes become spaced very close together. At low frequencies, the distribution of modes with respect to frequency determines the coloration of room response. In general, room dimensions that leads to an even and uniform spacing of room modes leads to the most natural sounding environment. Room dimensions that lead to common modal frequencies should be avoided since increased response at these coincident frequencies may cause irregular boominess. Similarly, a situation with large spacings (approximately 20 Hz or larger) between adjacent modes will also result in unnatural response.

Rooms where ratios of length, width and height are related by small whole numbers should be avoided since this leads to coincident modes and consequent boominess at these frequencies. A room in the shape of a cube would be the worst case since axial modes associated with all three dimensions of the room would overlap. Beyond these simple guidelines, more concrete design criteria become complex. The search for optimum room ratios has been going on since the time of the 1940s. Recent trends suggest the use of a geometric series which closes upon an even number when continued past three dimensions. See Table 3.2-2, in which the first set is created from the square roots of successive numbers: $\sqrt{1}$, $\sqrt{2}$, $\sqrt{3}$, while the second set is the successive powers of the cube root of two: $\sqrt[3]{2^0}$, $\sqrt[3]{2^1}$, $\sqrt[3]{2^2}$. For rooms in which one or more dimensions must exceed twice the smallest dimension, even multiples of any of the preferred ratios can be used. Set 3 utilizes the ratios of set 2, but with the length and width each multiplied by 2.

Room mode analysis and optimization is particularly important in small rooms such as studios, control rooms, announce booths and so forth. A look at the example in Figure 3.2-4 shows why. Small rooms have few modes at the lowest audio frequencies and the spacing between them may be excessively large. The lowest mode in a room is equal to $565/L$ where L is the longest dimension of the room. This is $\lambda/2$ for a sound speed of 1130 ft/sec. Audio energy below this frequency will not be supported at all by resonance and room response at these frequencies will be attenuated.

As stated previously, three types of room modes exist. While all three are significant, a reasonable first-order design and /or analysis can be made using only the axial modes. The tangential and oblique modes involve four or six room surfaces respectively and are likely to suffer more attenuation due to the absorption of the surfaces and greater likelihood of meeting physical obstructions. In general, an optimized axial mode

Table 3.2-2
Recommended room dimension ratios for optimum mode spacing.

	Height	Width	Length
Set 1	1.00	1.41	1.73
Set 2	1.00	1.26	1.59
Set 3	1.00	1.52	3.18

SPL Attenuation as a Function of Distance

In a free field, sound intensity varies inversely with the square of the distance from the source. Hence, like light, sound follows so-called inverse square law. The difference in dB between the SPL at two different distances, d_1 and d_2, is then:

$$\textit{Difference in } \mathrm{dB} = 20 \log\left(\frac{d_2}{d_1}\right)$$

This is equivalent to a 6 dB loss per doubling of distance or a 20 dB loss per tenfold increase of distance.

Directivity of Sound Sources

In the practical case, sound sources are not omnidirectional and have an axis of main radiation. This can be quantified in the concept of directivity, or Q. Directivity (Q) is the ratio of intensity along a given axis to the intensity which would be measured at the same distance if the same quantity of total acoustic power were being radiated omnidirectionally. The designated axis is usually taken to be the axis of maximum radiation, so $Q \geq 1$. This can be expressed as:

$$Q = \frac{(\textit{On-axis pressure})^2 \textit{ at some distance}}{(\textit{Mean sound pressure})^2 \textit{ at same distance averaged over all directions}}$$

It can be shown that for a sound source having theoretical horizontal and vertical coverage angles of a and b,

$$Q = \frac{180}{\sin^{-1}\left(\sin\frac{a}{2}\sin\frac{b}{2}\right)}$$

Directivity is often expressed in decibel notation:

$$\textit{Directivity factor}\ (\mathrm{DF}) = 10 \log Q$$

The attenuation of SPL as a function of distance given previously can be modified to include the factor of source directionality:

$$\mathrm{SPL} = \mathrm{PWL} + \mathrm{DF} - 10 \log (4\pi r^2)$$

where r is the distance from the source in meters.

Effect of Humidity on Sound Level

Humidity in the air will increase sound level loss in excess of that predicted by inverse square law alone. Figure 3.2-2 shows a family of curves at different frequencies showing dB loss per 100 ft as a function of RH. In general, the attenuation increases with frequency and absorption is greatest between approximately 10% to 40% RH, decreasing above and below these levels for all frequencies.

The curves show why it is difficult to acoustically transmit high frequency energy long distances. For example, the SPL at a distance of 4 ft required to produce 80 dB SPL at 1000 ft for 10 kHz at 20% RH is:

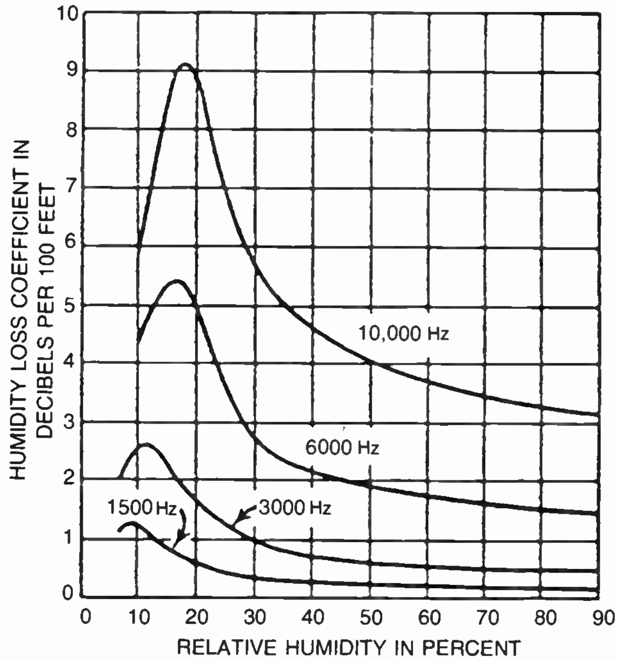


Figure 3.2-2. The effect of humidity on sound attenuation (dB loss per 100 ft).

$$\underbrace{80 + 20 \log\left(\frac{1000}{4}\right)}_{\textit{inverse square law term}} + \underbrace{9 \text{ dB}/100 \text{ ft} \times 10}_{\textit{humidity term}} = 218 \text{ dB SPL!}$$

SOUND IN ENCLOSED SPACES

Room Modes

It is well known that a guitar or violin string resonates at only selected discrete frequencies, determined by the length of the string and the tension on the string. It can be easily shown that these resonant frequencies include higher order harmonics, by placing a finger on the exact center of the string, damping out the fundamental frequency tone, but not the even harmonics, which continue to resonate.

Similarly, the resonant tone produced by an organ pipe is dependent upon the length of the tube, atmospheric pressure, and the end conditions; i.e., whether the pipe is open or closed. For a pipe closed at each end, the resonances are those of the standing waves in a closed tube, as shown in Figure 3.2-3. Once again, the resonant frequencies include the fundamental, for which the length of the pipe is one half the wavelength ($\lambda/2$) and other harmonics for which the length of the pipe is some multiple of one half of the wavelength (λ, $3\lambda/2$, 2λ, etc.)

Like closed tubes, rooms exhibit analogous characteristics of resonance . . . those frequencies at which sound is sustained or reinforced by the geometry and boundary conditions. However, the situation is now shown. As is evident, the air is alternately compressed and rarefied. In the second picture, at an instant of time later, it is apparent that the compression/rarefaction cycle is moving to the right.

The Speed of Sound

The speed of sound depends principally on the nature of the propagation medium in which it is traveling. Table 3.2-1 lists the speed of sound in various materials.

Looking at Table 3.2-1, it is easy to understand situations such as when banging on a pipe in the basement of a building results in the perception of a double bang in an upper floor room due to the arrival of sound energy transmitted via the pipe itself re-radiating in the room versus the slower airborne path through the room partitions and floor/ceiling structures. Also it is worth keeping in mind that loudspeakers closely coupled to a wall structure can in some cases transmit sound via the structure to be re-radiated from another point in the structure. Thus, since sound travels much faster in solid structures than in the air, sound energy can arrive at a listener or microphone before the direct sound path arrives, causing a weak but annoying pre-echo.

The speed of sound at normal room temperature is approximately 1130 ft/sec. However, the speed of sound is also dependent on temperature. At 32° F and 760 mm Hg the speed of sound has been experimentally verified to be approximately 1,087 ft per second. A reasonably accurate simplified formula for the speed of sound is to assume a 1.1 ft/sec increase or decrease for each degree F above or below 32°:

$$\textit{Speed of sound} = 1{,}087 + 1.1(T - 32)\ \text{ft/sec}$$

where T is temperature in degrees Fahrenheit

It is worth noting that the ambient air pressure is not a factor in the speed of sound.

Relative humidity changes the density of air and has a small effect on the speed of sound. At typical room temperatures, the percentage change in velocity between 0 and 100% RH is less than 0.5%. This factor is usually ignored in practical situations.

Definitions and Sound Measurements

A decibel (dB) is a mathematically convenient way for expressing the ratio of two power-like quantities:

$$\text{dB} = 10 \log \left(\frac{P_1}{P_2}\right)$$

where P_1 and P_2 are power quantities. Units expressed in decibels are designated as *levels* and are defined as follows:

$$\textit{Intensity level}\ \text{IL} = 10 \log \left(\frac{I}{I_{ref}}\right)$$

where

I = Sound intensity (watts/m^2)
$I_{ref} = 10^{-12}$ watts/m^2

$$\textit{Acoustic power level} = \text{PWL} = 10 \log \left(\frac{W}{W_{ref}}\right)$$

where

W = acoustic power (watts)
$W_{ref} = 10^{-12}$ watts

$$\textit{Sound pressure level} = \text{SPL} = 20 \log \left(\frac{p}{p_{ref}}\right)$$

where

p = Sound pressure (N/m^2)
$p_{ref} = 2 \times 10^{-5}$ N/m^2 or 0.0002 microbar

In a free field,

$$I = \frac{p^2}{\rho_o C}$$

where

ρ_o = Air density
C = Speed of sound
SPL $\approx$ IL in a free field.

In a diffuse field,

$$I = \frac{p^2}{4\rho_o C}$$

$$\text{SPL} = \text{PWL} - 10 \log (4\pi r^2) + .5$$

where

r = distance from the sound source in meters

and the source is omnidirectional.

When adding decibel quantities, it must be remembered that power levels themselves can be added directly but decibels cannot. To add decibel quantities, the decibel notation must be rearranged so that the associated power levels are added and then converted back to decibel form:

$$\text{dB}_{total} = 10 \log [10^{\text{dB}_1/10} + 10^{\text{dB}_2/10} + 10^{\text{dB}_n/10}]$$

For an approximate rule of thumb in adding decibels, refer to the following table:

When adding two dB values that differ by:	Add to the higher value to obtain the total:
0 or 1 dB	3 dB
2 or 3 dB	2 dB
4 to 8 dB	1 dB
9 or more dB	0 dB

Table 3.2-1
The speed of sound in different media.

Medium	Approximate Speed of Sound
Air (room temperature)	1,130 ft/sec
Soft Wood	11,000 ft/sec
Gypsum board	22,300 ft/sec
Steel	16,600 ft/sec
Water	4,900 ft/sec
Glass	17,000 ft/sec
Concrete	11,200 ft/sec
Hydrogen	4,100 ft/sec
Helium	3,000 ft/sec

3.2
PRINCIPLES OF ACOUSTICS FOR BROADCAST APPLICATIONS

LYNN D. CLAUDY*
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC

INTRODUCTION

Acoustics—*the scientific study of sound, especially of its generation, propagation, perception and interaction with materials and other forms of radiation.*—American Heritage Dictionary.

Reducing the above definition to its basic core, acoustics encompasses essentially *the total effect of sound*, including the characteristics of the sound source, propagation medium and the reception of the transmitted sound. This chapter covers basic principles of acoustics relevant for broadcast applications. In addressing this subject, a review of the basics of sound wave propagation will first be presented. The behavior of sound in enclosed spaces will then be examined and design techniques for controlling sound within a room will be introduced. Sound isolation techniques, or controlling the intrusion of unwanted noises into enclosed spaces, will then be presented. The chapter offers a broad overview of acoustics, with a blend of theory and practice, with the goal of providing a balanced foundation on the subject. However, as the practice of acoustical design is still part science and part art form, the services of a professional consultant in this area will continue to be just as appropriate for the knowledgeable reader as for the layman.

THE NEED FOR ACOUSTICAL DESIGN IN BROADCASTING

The importance of the need for good acoustics in broadcasting applications cannot be overstated. Borrowing the garbage-in-garbage-out metaphor from the computing industry, it is easy to imagine that an excellent transmission system can be a wasted resource if the signal source itself is degraded. High ambient noise, improper acoustical treatment or poor noise isolation can quickly alienate, fatigue or annoy a listener. Common sense might lead one to a different conclusion: often one experiences poor acoustic environments in daily life and is not overly adversely affected: conversation does take place in noisy corridors, musical performances are well received in compromised settings, and so forth. Noise, excess reverberation or extraneous sounds are typically much less irritating to listeners when presented contextually in their original environments than when reproduced over radio or television. The human binaural sense of hearing, visual and other cues allow concentration on a desired source and inhibition of unwanted sounds. A simple experiment can be set up to prove this point. Carry a pocket tape recorder to a cocktail party and record the environment while moving from one conversation circle to another. While it is relatively easy to understand live conversations at the party, the recording will in all likelihood be a cacophony of jumbled sounds, indistinguishable one from another. The lesson: acoustics, while being of significant importance in live settings, is ever more so important for radio and television broadcasting because any defects that are present are magnified when the sounds are taken out of the originating environment and reproduced in another.

*Additional material provided by Eric Neil Angevine, Oklahoma State University, Stillwater, OK.

BASIC CHARACTERISTICS OF SOUND PROPAGATION

Sound is a wave phenomenon. The wave motion is of the longitudinal (as opposed to transverse) variety: the air molecules move back and forth in the same direction as the wave motion. All molecular motion occurs with respect to any motion of the air; there is no net motion of the air particles within the air. Figure 3.2-1 shows a sound wave traveling down a tube at two points in time: the distribution of air particles and regions of high and low pressure in the tube are also

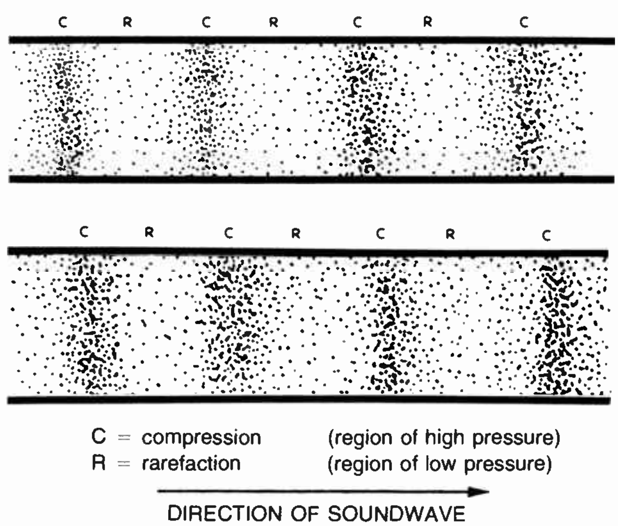


Figure 3.2-1. Illustration of sound wave propagation in a tube.

Doors to studios and control rooms should be of solid construction with weatherstripping (felt, not rubber that will squeak) and a good quality door closer. Do not use a latch on the door that will click each time the door is opened and closed. Doors should not open out into busy hallways where people will walk into them as they pass by. The average door in most stations is 3 ft wide. That can be reduced by 6 in. However, most furniture requires up to 30 in. to be moved through a doorway and more if it opens into a narrow corridor. The narrowest hallways should be no less than 3 ft 6 in. Note that hallways take up space. Every square foot of hallway that can be eliminated yields an equal amount of space that can be used for something productive.

Where the location of the facility allows, windows to the outside should be considered. This is particularly applicable for radio stations. In some instances studios are built in the interior of a building and exterior windows are not possible. However, when a radio studio or combo facility is built on an outside wall, a small exterior window will allow occupants to know a little more about the weather and have a connection with the outside world. Of course, if the facility is located next to an airport or major highway, the window invites more background noise than might be acceptable.

Sound locks, while desirable, are not essential, but can be combined with other purposes. Plan on plenty of lights but place the ballasts of fluorescent lights out of the production areas to eliminate acoustic noise and interference.

If the transmitter is at the studio site, locate it and the main control room so that meters on the transmitter and its associated monitoring equipment can easily be seen from the operating position.

Equipment

In a low-budget situation remember this rule of thumb: *Will the equipment to be bought pay for itself and make money for the station?* The purchase of a \$200 directional microphone to do a one-time remote broadcast that pays only \$100 is poor business.

Use this logic when equipping the station. Create a list of those functions you will have to perform and spend only what is needed to get on the air and function. However, plan for the future. Design around a console with enough inputs to accommodate those unanticipated needs. A four potentiometer board is not enough for most on-air operations except for simple production. Even the smallest stations should have provision for six or eight simultaneous inputs. Invest in a good quality telephone interface system. Virtually all stations now take phone calls for direct on-air use.

Consider used equipment with great care. While older equipment was built for long life, obtaining spare parts may be a problem and the general quality level may not meet today's needs.

Consider versatility and contingency in equipment configurations. The studio output can be operated through the automation system as a source. In the event of an emergency, the announcer can easily connect the output of the console to the studio-to-transmitter link (STL), bypassing the automation system. In the event of an STL failure, the audio can be patched directly to the transmitter via the remote control system line or a telephone line.

Do not forget the patch panel in the control room. There is a growing tendency among engineers today to eliminate this important switching center. All high level inputs and outputs should appear here. Even console inputs should show up. This also applies to all recorders, tuners and other audio sources. This practice makes for a very versatile operation and can save the embarrassment of dead air time if the console should fail. If this happens, a player can be patched directly to the transmitter, bypassing the defective elements in the system.

Consider every piece of equipment in the audio chain as a potential single point failure problem and address a plan on how to bypass it or do without while it is being repaired. Consider also a backup power generator. For a small station a modest sized generator can power the entire facility except for a central air conditioning system.

The construction of a new or renovated facility is a good time to implement a checklist approach to routine testing of all equipment on an annual, quarterly, monthly or weekly basis depending upon how often the equipment needs to be checked. The checklist should also include such items as the air conditioning and heating system, backup generator, tower, antenna, lights and the general condition of the facility and surrounding environment.

CONCLUSION

The planning of an audio production center is arduous at best. Every project reaches a point where the deadline is met and the budget must be closed. However, when everything is up and running the job may still not be done. It is not unusual to see 85% completion at this point. Those who control the purse strings will ask *we're in production, why go any further?*

The sheer exhilaration of a new facility will give the production and operations people the energy to carry on. When it wears off the end-users will tire of having to put up with inadequacies every working day. The product invariably suffers from lack of efficiency and enthusiasm.

The construction group will have moved on so the technical maintenance team is likely to handle the follow-up tasks. Because this must be done in addition to routine activities, budget some overtime for this purpose. The process will take three to six months. Keep communications open with the users. They will be patient if they see progress and know that someone is sweeping up corners and tying up loose ends.

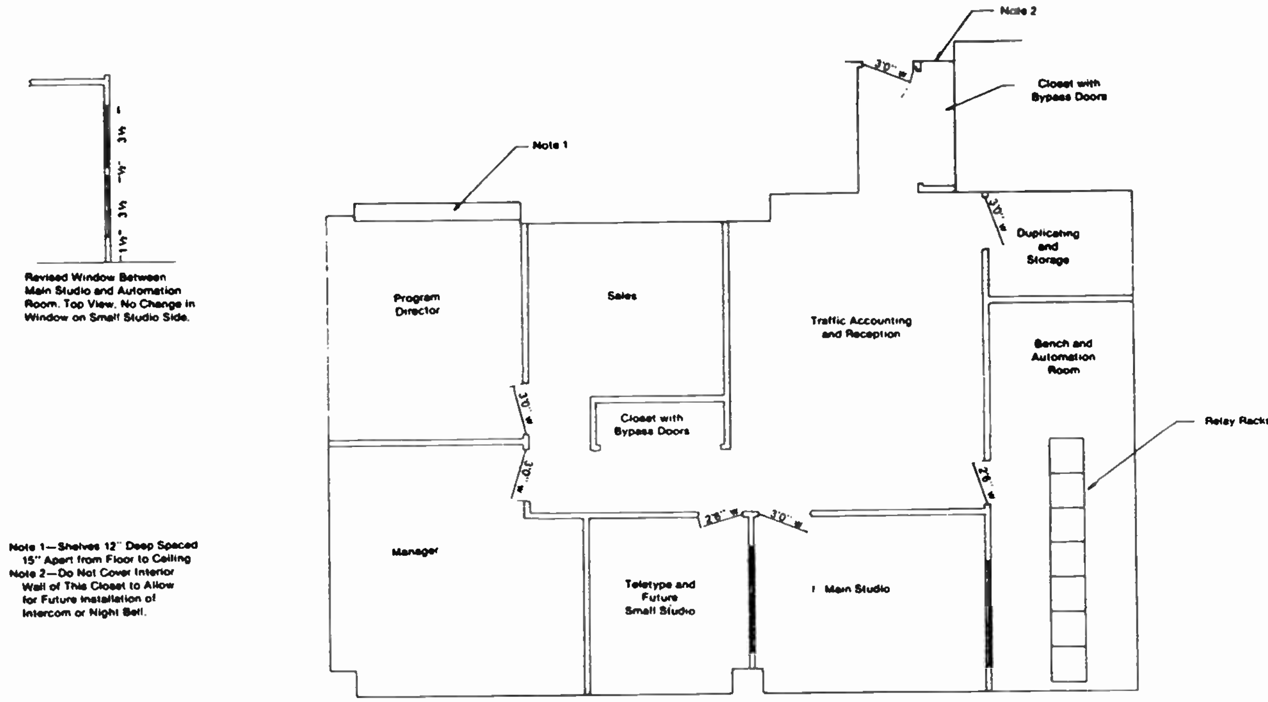


Figure 3.1-11. Studio layout for a small station.

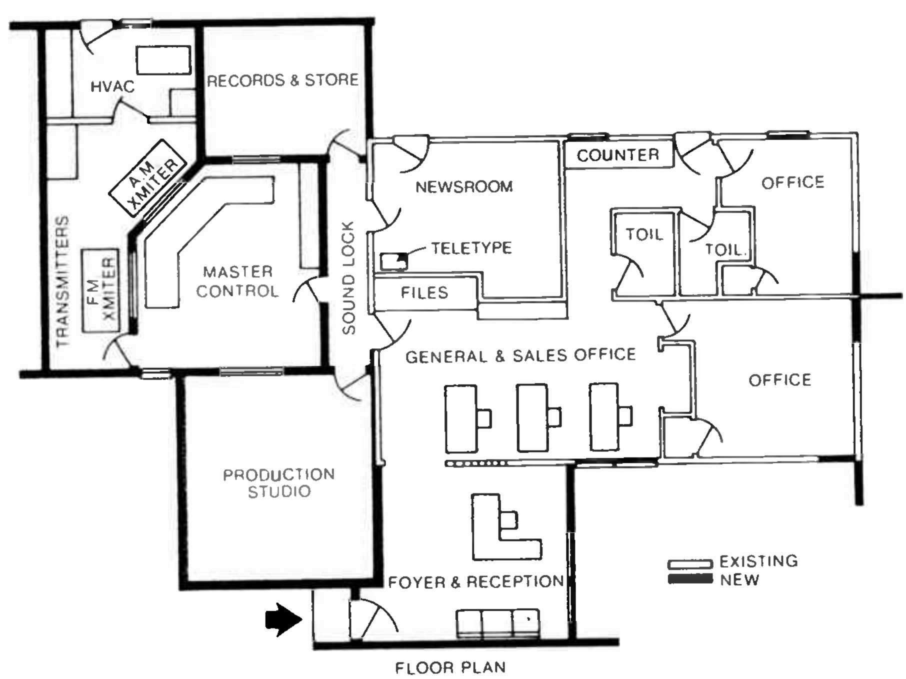


Figure 3.1-12. Renovated small station.

material is reinforced concrete. A wooden floor can be used if it is reinforced to prevent bouncing. Otherwise people walking on such a floor will cause problems with turntables and compact disc players. Sandbags or bricks in the base of the turntable cabinet will sometimes improve an otherwise bad situation.

Avoid placing the console up against a wall. In addition to being an acoustic problem, the operator will find looking at a wall to be uncomfortable. Locate the console in the middle of the room as a desk might be in an office. The wall space can be used for shelves, equipment, or other facilities.

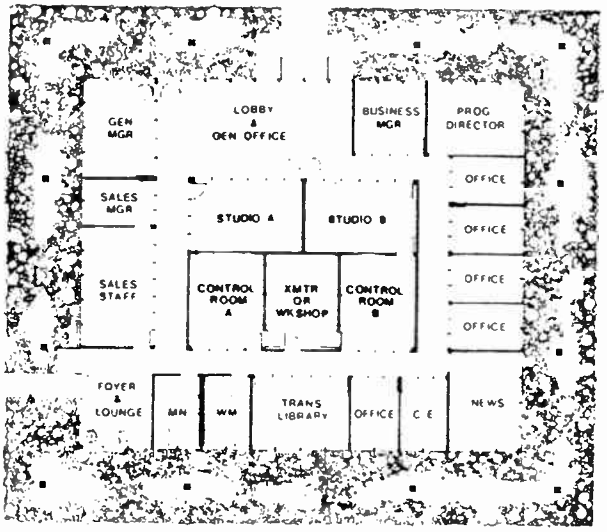


Figure 3.1-10. Large station. Metropolitan market full staff.

station's announcing is carried on and may be the only studio area in a station other than the production studio (also a combo). Combos usually contain most of the audio equipment of the station. Frequently, a transmitter and record library will also be located in the control room.

In many instances the news room will be used with live mics for late breaking news events. Both situations make the problem of acoustics more complex but, because radio sells with sound, good acoustic design is essential when the best sound possible is required.

Design factors of combo control rooms are at best a compromise of the several design factors:

- Location of the control room within the studio building
- Isolation. Elimination or substantial reduction of unwanted sound and noise, both internal and external
- Construction. Special walls, ceilings and floors may be required
- Reverberation control. Size the rooms and consider special wall and floor coverings to eliminate or reduce reflections and resonances that degrade the studio sound on the air
- Ventilation and air conditioning. Technical equipment generates more heat than normal office equipment and studio guests add to the heat load that must be accommodated with the HVAC facilities
- Size and arrangement of equipment. There are as many arrangements as there are designers—plan some flexibility for changes that will be needed after the facility is used for a while and for accommodating future needs.

RADIO STUDIOS ON A LOW BUDGET

Considerations Involved

Different sized stations require different construction criteria. A 50,000 W AM clear channel or Class C FM in a major market and a 250 W AM daytime or Class A FM in a rural area dictate that the small market station is going to be constructed quite different from its big city counterpart. In most cases, the selection of a studio site in a small market is dictated by what it costs to get the space. It is not unusual for space to be traded out in part or in full for advertising. Figures 3.1-11 and 3.1-12 show examples of small radio broadcast studio and office facilities.

For the construction or renovation of a smaller facility, the chief engineer (or engineering consultant) may be presented with an existing suite of offices, a store front or even an old house that must be converted into studios and control rooms. After determining the nature of programming management intends to offer the community, the facility can be designed to meet the requirements. Most smaller stations use combo control rooms and studios, a simple production facility (that can be used as an emergency control room) and a newsroom with recording and editing facilities capable of on-air use.

A larger studio can be fashioned from the lobby area, the general office area or by combining two or more management offices. With some advance planning, these areas can be acoustically treated and wired for use as acceptable studios for those occasional talent shows, church programs, talk shows or special group events that do not occur often enough to justify having a full-featured studio available. If the station is a new or growing operation and cannot afford to do all of the above, at least consider what functions may be required at a foreseeable future date. Plan now for what might be needed in two or three years.

Physical Layout

While the total amount of space may be limited, remember that combo control rooms are occupied for long periods by one or at most two people. Consider this as their office that should be large enough to house the equipment, move around, have visual access to other areas (including the outside) and be comfortable and well lit.

When laying out the studio location, draw a sketch of the available floor space and existing walls. Make a number of copies of this floor plan and start drawing in the studios, offices, newsroom, reception areas and engineering. Make three or four versions of the plan. If the existing layout will allow the installation of new walls, avoid perfectly square rooms. A room that is 8 ft × 8 ft × 8 ft is going to have a very definite and unpleasant resonance. Strive for 6 ft × 8 ft × 10 ft or dimensions in this proportion. Be sure to add extra wall thickness for the studios and control rooms.

Consider laying out the facility using scale cutouts of typical office furniture, studio equipment and consoles, plus storage cabinets and counter tops. Add some closets and undefined storage spaces for the future. Keep the main studio and control room from beneath heavily traveled stairways and corridors, air conditioners and front windows opening on busy streets. The control room floor should be very solid. The most desirable

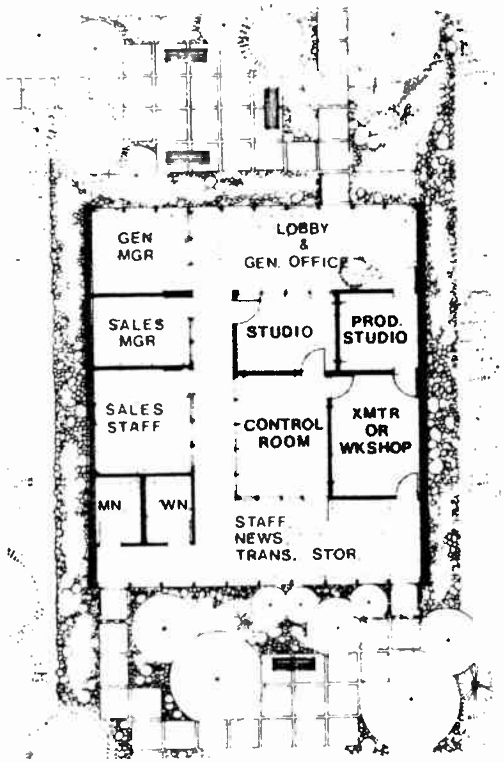


Figure 3.1-7. Small size AM station.

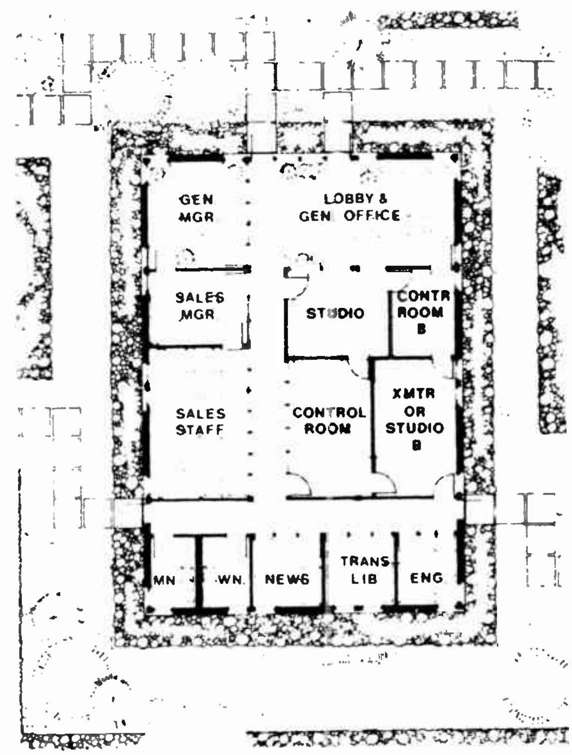


Figure 3.1-8. Medium size AM station.

ated studio (see Figure 3.1-9). Control rooms are separated by the transmitter, automation or engineering area. This floor plan includes approximately 3,150 sq. ft and is suggested for stations planning both AM and FM operations. Additional office space is also allocated for the larger staff in this station. This design features a mirror glass curtain wall building set in a reflecting pool. The building costs approximately $90,000 to $135,000.

Plan Four. Large Station

This 4,300 sq. ft studio/office complex is an impressive broadcast center (see Figure 3.1-10). Of primary importance is the location of all control room and studio space in the center of the building, eliminating the problem of outside traffic noise in a metropolitan area.

Operating personnel are assigned to the rear office areas and the news room is strategically located near the control rooms and an outside exit to the news cars and trucks.

The building is made of exposed concrete and features a dark glass curtain wall. The building cost is approximately $100,000 to $150,000.

CONTROL ROOM/STUDIO DESIGN

Many present day radio control rooms also double as studios and are called *combos* for combination control and studio. The combo room is where most of the

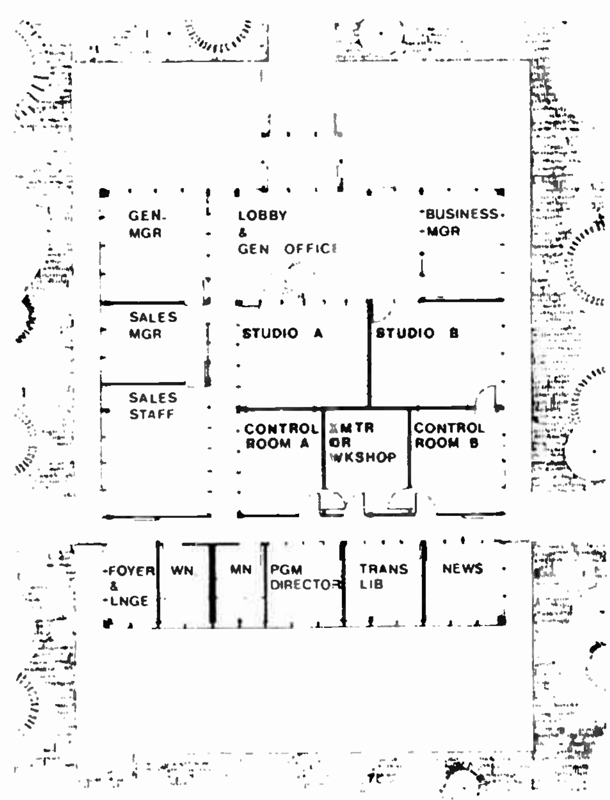


Figure 3.1-9. Large station AM/FM dual control room.

If 1/3 octave graphic EQ must be used, use smooth curves with no more than 2 dB difference between any two bands. Leave it bypassed for a few days, then switch it in to decide if you really like what it does.

STORAGE SYSTEMS

Professional CD, DAT and digital cart machines are the plow horses of most production centers. High quality R-DAT units offer economical multitrack operation for off-line production. Third generation computer based virtual multitrack systems are a good alternative but present the same DSP hazards as digital mixers.

Central digital storage systems are coming of age. They offer the advantage of simultaneous access to all material from every work station. These systems are worth considering for their ability to save large amounts of storage space. Redundancy is a given but no on-line operation should give up its traditional machines and one closet full of materials to play in them. Recordable CDRs offer a convenient backup of archival materials.

Processors

Every audio production benefits from a collection of processors. Tools that compress, limit, reduce noise and remove hum are useful in conditioning materials for the best sound. As with microphones, look at what others are using. Remember that newer gadgets may offer no real advantage over some of the old favorites.

Effects Devices

These are used to enhance the texture of sound or change its character. Reverb, delay, phasing, flanging and the like are enhancements. Pitch shifting often is used to disguise voices. Newer devices are highly sophisticated and may be able to do all of these effects. If more than one effect is needed on the same sound, multiple units may be required. Installations with multiple control rooms have the option of either moving effects gear around as needed or using tie trunks to access unused gear in other rooms.

ESTIMATING COSTS

Many of the factors that affect the costs of acoustic construction are actually under the station's control. A top 40 radio station whose talent works only a few inches from a single microphone, relies less on room acoustics for its live sound than does an all-talk format station requiring several open microphones to pickup guests at a variety of distances. Because broadcasting is considered high fidelity, studio acoustics must be closely controlled.

The more audio signal compression and limiting a station employs, the more evident will be the annoyance of poor studio acoustic characteristics. Conversely, the compression on stations with better studio acoustics will be less noticeable during live broadcasts and recordings.

AM AND FM STUDIO FACILITIES

Building Planning

One of the prime requisites for a successful broadcast station is the careful layout of studio, production and administrative areas to achieve maximum effectiveness of space and personnel. The following four typical layouts depict a range of facilities from a small market minimum staff facility to an arrangement suitable for a large metropolitan operation employing a full complement of personnel. Each floor plan is handled differently according to the needs of different size stations.

In these examples, the control room, studio and production facilities are in a centrally located core area although they need not necessarily be there. The suggested sizes of these areas should be considered as minimum from an operating standpoint with normal equipment complement. The layouts are presented as a guide for planning a modern, functional radio facility with considerations given to size of market, staff and programming requirements.

Plan One. Small Radio Station

With approximately 1,800 sq. ft, this floor plan provides adequate space for the small AM or FM station with a minimum staff (see Figure 3.1-7). Since smaller staffs have several responsibilities, partitioned general office space is omitted in favor of a large news, recording storage and general use areas at the rear of the building.

The transmitter and engineering workshop area is placed next to the control room with a window recommended for a clear view of the transmitter meters.

The building is of brick and plaster fascia and includes a glass curtain wall. The building price will vary considerably depending on area construction costs; but a typical range is $60,000 to $75,000.

Plan Two. Medium Radio Station

In medium-sized stations, office space is increased for sales, promotion and programming activities but without a substantially larger technical core area. This floor plan takes the small station layout and expands it to approximately 2,500 sq. ft (see Figure 3.1-8). More room is provided for the sales staff and clerical help and an impressive office for the general manager. Studio and control room space is slightly larger in anticipation of more equipment and activities in these areas. The news director, recording library and chief engineer gain office space. Alternate core area layouts may be employed.

The building includes brick walls, weathering steel columns and fascia with a dark glass entrance and glazing strips. Cost is approximately $65,000 to $90,000 but may vary considerably, depending on construction costs in a particular area.

Plan Three. Larger Radio Station

In this plan, the technical core area is adequate for two full-size control rooms, each with a large associto see microphones and processors dating from the 1950s and 1960s in photos of recording studios. The old favorites are still around because they are perceived to have personalities unmatched by newer choices. Surely this is true but as progress continues to peel back the veils of distortion, some of those characteristics will go out of favor.

Microphones

The electromechanical nature of microphones contributes to their imperfection. Each of the three basic types—dynamic units, condensers and ribbon microphones—has merits and disadvantages. Dynamic units are rugged and the best examples offer very smooth, wide frequency response. No power source is needed and their domed diaphragms diffuse reflected sound waves. Thus far, dynamics cannot be made as small and lightweight as required for body microphone uses.

Condenser (capacitor) microphones provide higher output, overcoming electronic noise but require a power source. Their necessarily flat diaphragm reflects arriving wave fronts resulting in cancellations. Large diaphragm condensers also create acoustical phase distortion as wave fronts arrive at different times across the element. This may be perceived as a slight harshness in the higher frequencies. Very sibilant voices risk digital overload from HF peakiness. Tiny clip-on condensers approach the theoretical perfection of a point in space.

Perhaps the best because of their ultra-lightweight element are ribbon microphones. Unfortunately they are extremely fragile and prone to damage from a mild breeze. Time will tell which type really is best. Find out what others are using and try one of these microphones first. When seeking a new microphone, look for units that sound good with minimum equalization (see also Chapter 3.3 "Microphones.")

Mixers

In addition to the human factors already discussed, the choice of a mixer must consider sonic quality. Low noise, good fidelity and resistance to interference are expected. Simple three-band quasi-parametric equalizers offer more than sufficient control in the hands of a knowledgeable user. Four-band EQ is more susceptible to abuse, causing irreversible phase shifts. Onboard gating and compression may be adequate. However, many users choose outboard equipment for these functions.

The dust cloud surrounding digital signal processing (DSP) is continuing to settle. Manufacturers and users are beginning to understand the subtle artifacts introduced by insufficient digital word size and improper dithering (rounding off of digital calculations). Even the ultimate high-end digital systems may still have some of these problems. Until everything is fully digital there will continue to be good arguments for a high quality hybrid system with pass-through digital inputs and both types of output.

Loudspeakers

In a production room the monitoring loudspeakers must be as accurate as possible. Too often program is over-equalized to make up for monitor problems. It is not sufficient to select a system because it is popular and found in similar facilities. Instead it is necessary to perform an objective evaluation of the potential choices and select one that is unquestionably the most accurate.

Evaluation of Loudspeakers

A/X/Y testing of loudspeakers maintains objectivity and overcomes the variables of other methods. *A* is a known, *X* and *Y* are unknowns to be compared to A and to each other. The testing may be done in a large room, parking lot or open field. If an existing control room is available, listening should be done there. Otherwise a small mixer, headphones and a few power amplifiers may be used.

The process consists of placing two candidate systems together and laying a microphone on the floor or other hard surface about 15 ft away. An omnidirectional dynamic microphone with wide, smooth frequency response is recommended.

Using program material containing music and speech, setup the mixer to compare direct program with that returning from the microphone. Switch repeatedly between the direct and each of the test loudspeakers. The more accurate one will reveal itself easily; replace the other with the next test unit. To ease any doubts about the control room monitor, use a similar monitor as a test system. Any inaccuracies will be magnified, not masked. Finally, rerun the test using the initial winner as a control room reference to check the top two or three candidates. The results should not vary significantly.

The theoretically best loudspeakers are coaxial in construction. Next are those with drivers aligned in a vertical array. There should be only one driver for each frequency range; two-way systems tend to be more accurate than those with three drivers. This is due to inevitable distortions in the crossover range. Systems with drivers horizontally offset create acoustical phase distortion. This is due to different arrival times, at the ears, of sounds in the crossover band. Loudspeakers built in mirror image pairs are the first to delete from the list.

Installation

Follow the manufacturer's recommendations for installation. Most systems are intended for 1/2-space mounting against or flush within a wall. Some are meant for 1/4-space locations such as wall-ceiling corners. Placement away from these solid surfaces results in reflections that damage low frequency performance.

Human hearing readily conditions to the nature of sound in a room. A friend's voice is as easily recognized in the kitchen as in the living room. Equalizers for voicing a monitor system may add more phase distortion than the response errors they correct. Confidence in the monitor's accuracy usually is sufficient.

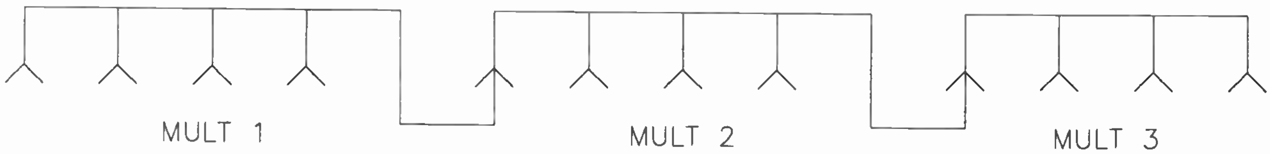


Figure 3.1-4. Half-normaled mult groups.

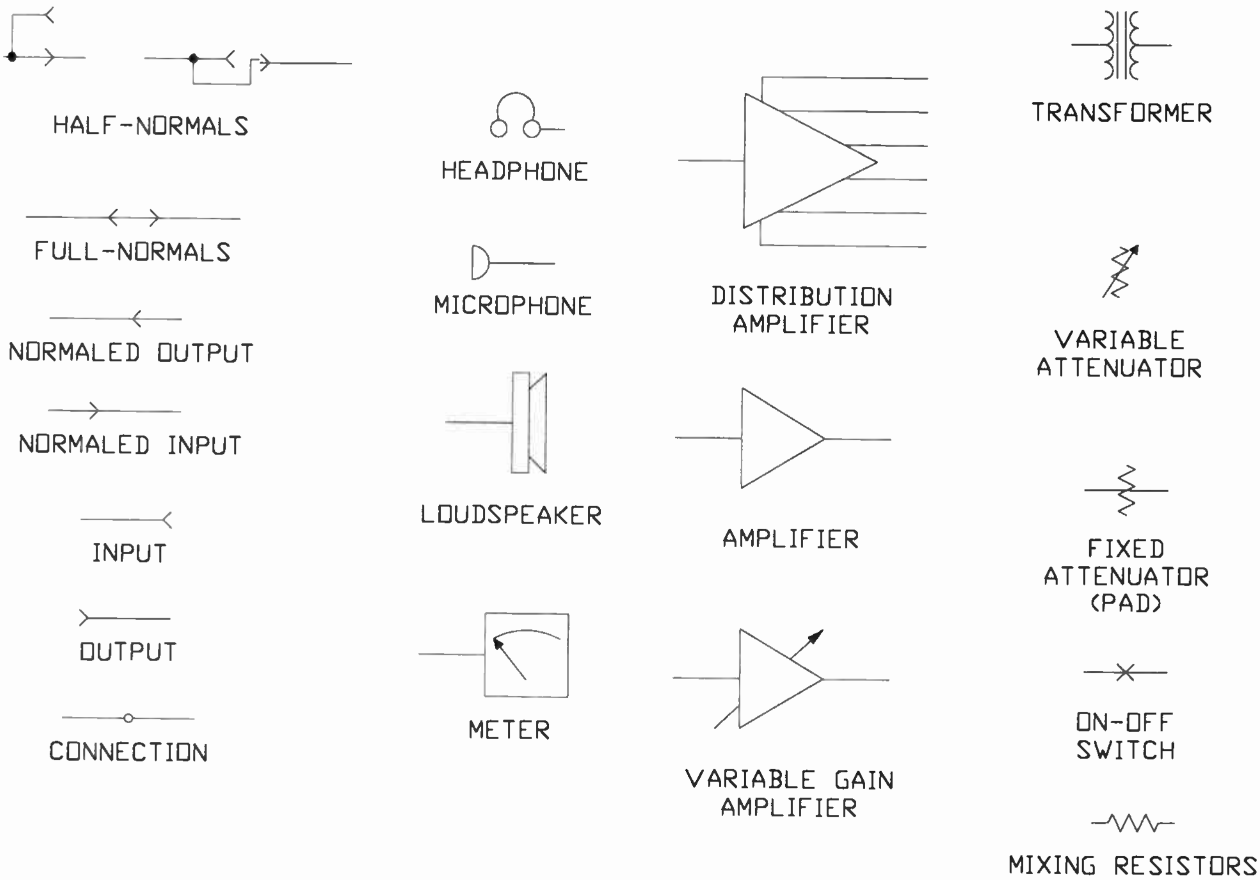


Figure 3.1-5. Common symbols for audio equipment.

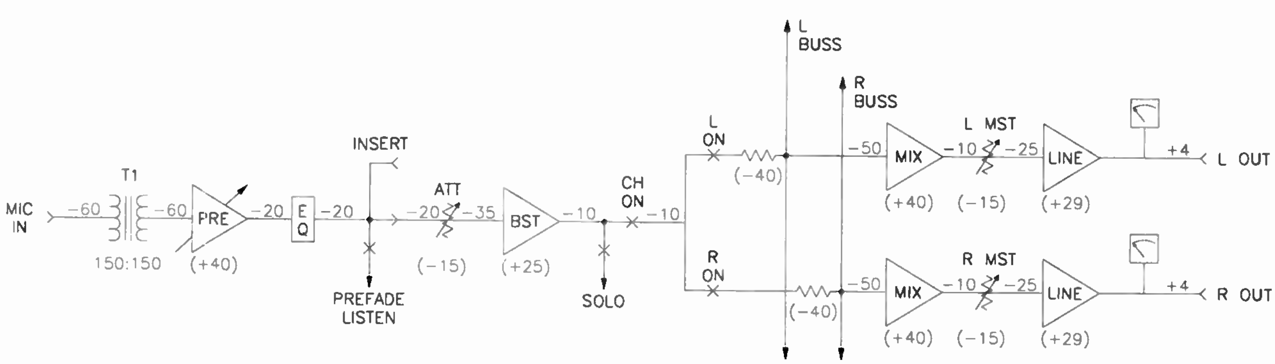


Figure 3.1-6. Single-line of typical channel path.

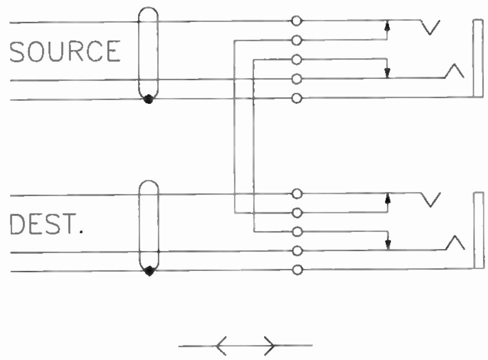


Figure 3.1-1. Schematic and symbol for full-normaled jacks.

plug in the destination jack breaks the path and injects a new source. The break jack may be identified by a dot in the lower left corner of its label. Of course, operators must be told of the dot's meaning.

A third configuration is called *blind normals*. Here the source is not available but a different source may be sent to the destination if necessary. This is useful when sufficient DA outputs are available and patchbay space is becoming limited. Figure 3.1-3 shows the scheme for using a blind normal for transmission. This method is preferred over half-normals because it avoids introducing a possible short circuit across a critical signal path.

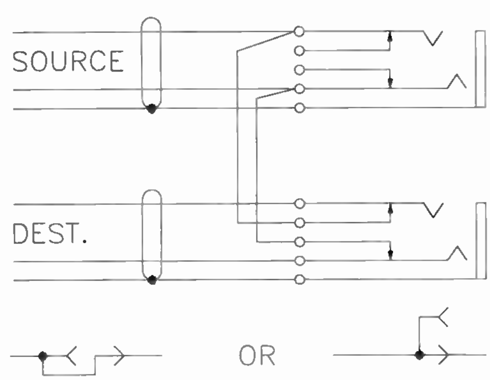


Figure 3.1-2. Schematic and symbols for half-normaled jacks.

Patchbay Utilities

Patchbay layout takes a special knack. Things that are alike should be kept together and multiple sources to a single destination should be grouped by destination not source. Some color-coding is helpful but not at the expense of legibility. White printing on bold colors is best because the eye irises to the white information rather than the darker background.

There are some less used utility devices and circuits that bring cheer to operators when they are needed:

- External device circuits connect to panels of XLR connectors near where a piece of equipment is likely to rest for temporary use. Six males and six females on each side of the room are a good minimum.
- Three-position stereo utility keys (4)
- Utility 600:600 transformers (4)
- Utility variable attenuators (4)
- Utility fixed attenuators (4 full-normaled sets—10 dB + 15 dB + 20 dB)

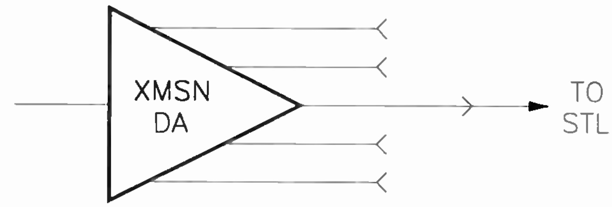


Figure 3.1-3. Symbol for using blind normals for transmission.

When all of the regular circuits are accounted for there will be gaps where some jacks have no assignments. The scattered sets of three or more jacks are suitable for creating *mult* (for multiple) *groups*. These are jacks wired in parallel and are useful for splitting a source to more than one destination. Although DAs are now more common for this purpose, some mult groups will always come in handy. An added convenience is obtained when the groups are consecutively numbered and half-normaled together as shown in Figure 3.1-4. This increases the available outputs without using cords to tie the groups together. Do not wire the last group back to the first one.

The Single-Line Diagram

When the facility is completed there will be many pages of detailed installation drawings. These will reside in the files of the technical maintenance shop. Each production room should have a simplified road map to help the operators find their way through the system. The audio single-line diagram minimizes the details in favor of clarity. Rather than every wire and connection, one line is used to represent a path.

A typical input channel will be shown along with master channel routing through the transmission DAs. Ancillary production equipment in the system should also be shown. The utility circuits and devices, along with their patch bay locations are also included. Figure 3.1-5 contains the symbols commonly used to abbreviate the components. Figure 3.1-6 is a very simple path from an input to the mixer's output. It shows how the gains, losses and nominal levels are indicated.

For a larger system it may take several hours to grasp the many paths and capabilities. For this reason it is helpful to find some convenient wall space to hang a copy of the single-line diagram. Remember it will become part of the room's acoustics. Sometimes there is a wall behind the equipment racks where such a reflector will do little harm.

PRODUCTION EQUIPMENT

Through the life of a production center there will always be a wide age range of equipment. It is common level. A pullout shelf at script apron height provides a place to put materials before and after use. The lower part of the rack may contain patch bays and distribution amplifiers (DAs). Computer or video monitors should be centered at or just below eye level. Continuous viewing of an object more than 15° above eye level can lead to nerve and muscle damage in the neck and shoulders.

Systems formed of conventional metal racks and cabinets are highly functional but somewhat sterile and mechanical in their appearance. Custom cabinetry using wood accents and plastic laminates is attractive but can be inflexible. Seek to use a modular approach with separate cabinets that fit together to form an integrated appearance. This will simplify changes of major elements as the installation evolves with production requirements and technology.

Communications

Good visual and verbal communications may be taken for granted but are not automatically inherent in a production facility. Some care should be taken to ensure open sight lines between the participants. A nod, wink or gesture might be critical to a project's success. The window between adjoining rooms must be wider and taller than normal. This is because the parties might be seated or standing and groups may spread out for comfort and ease of interaction. The basic assumption is that almost every place in each room should be visible to the other.

Voice communications range from the simple talkback built into most mixers to sophisticated multistation intercoms tailored to larger facilities. Almost every production center benefits from some form of modern matrix intercom with built-in *interrupted feed back* (IFB) and telephone interface. The advantage of a central matrix is that all users can be interconnected as production activities require. An important feature is separate volume controls on the master stations for each incoming channel.

AUDIO INFRASTRUCTURE

When the number of rooms in an audio production center exceeds two, a *master control* area is recommended. This creates a central place for access to the circuits that tie the rooms together. It offers a location for the intercom matrix and central digital storage systems. Sometimes called *central switching*, master control manages the incoming and outgoing broadcast feeds.

Cable Ways

Several miles of cable may be hidden away out of sight. Installations built with *computer floors* benefit from easy access to sub-floor cable raceways. However, a computer floor in an audio control room can be a disadvantage. The floor is resonant and sound isolation from adjacent rooms is more difficult. Building corridors on computer floors is one solution but few facilities have the luxury of combining solid floors with computer floors.

Suspended ceilings are a good alternative. A network of overhead raceways can follow the corridors, dropping down in closet spaces for passage into the rooms. Multiple 4 in. diameter thin wall conduits should be provided, not only for isolation of the various audio levels, but to ensure ample room for expansion. The traditional way to maintain acoustic integrity is to stuff the ends of the conduit with cloth. Remember to leave a nylon pull-rope in each conduit.

The following types of circuit should be kept separate:

- Microphone level audio
- Line level audio
- Digital audio and timecode
- ac power
- Loudspeaker level audio

Inside the control rooms a raised floor at the sides and behind the equipment provides a tidy way to route the cabling. Carpeted sections of 3/4 in. plywood resting on 4 in. × 4 in. boards make it easy to access the wiring when changes are needed. Whenever any set of circuits is permanently abandoned, remove the old cables to keep the area as open and clean as possible.

Patchbays

Practice has shown that eventually some operation will arise that exceeds the number of tie lines available. Amortized over the life of the system the cost of 24 rather than 12 trunks from every control room to master control is minimal. Add to these 12 AES/EBU digital trunks. (As digital audio settles into the broadcast world, they will become useful.) These are just the spares, not used for the normal operation. In addition, there will be the regular studio outputs, incoming sources from a central routing switcher and possibly a central digital storage vault.

Within the control room there should be patch bay access to every possible circuit:

- Incoming microphone circuits
- Line level sources
- Mixer inputs, inserts and outputs
- Machine inputs and outputs
- DA inputs and multiple outputs
- In-room trunks to a panel on the rear wall
- IFB program inputs
- IFB outputs
- Power amplifier inputs and outputs
- Loudspeaker inputs

Normaling Practices

Source and destination pairs of jacks that are wired together are called *normaled* because no patch cord is needed for their normal signal flow. Figure 3.1-1 shows the wiring for *full-normals* and indicates how a plug in either jack interrupts the circuit.

The impedance characteristics of contemporary audio devices have brought a newer practice called *half-normals* or *monitor* and *break* jacks. Figure 3.1-2 shows that a plug in the source jack does not break the circuit, but instead monitors the signal. A the control room. Here, extra care must be taken for acoustic isolation.

A good approach to layout places the sonically sensitive rooms in a central core surrounded by offices and other support areas. These serve as acoustic buffers against noise from outside the building. Production rooms may be thought of as islands in the core separated by corridors, equipment rooms and storage areas. It is helpful to provide access to control rooms from both sides of the core but not to the exclusion of a corridor across the core.

Traffic Patterns

Give careful attention to traffic patterns when determining where surrounding offices are located—the most used paths should be the shortest. Back-to-back control rooms may have doors to the perimeter corridors but this can lead to staff taking shortcuts instead of using the crossover corridor between them. If the control room doors are off of the crossover, they should not be opposite each other. Instead they should be offset by at least the width of one doorway.

A curious aspect of human nature is to congregate in the middle of high traffic areas. Two people will stand just outside a control room door and talk, with apparent disregard for those who would enter or leave the room until someone tries to get by. The intersection of two corridors is the same. Small groups will congregate in the middle of the crossroads, perhaps because the space seems large enough. Attempts to provide convenient alcoves for this purpose usually fail. The best solution is to make the corridors wider. Two or three people occupy a 4 ft diameter circle suggesting 6 ft 6 in. as a minimum corridor width.

HUMAN FACTORS

Ergonomics is the science of designing and arranging things people use for safe and efficient interaction. Each device in an audio system has been given some ergonomic consideration. When these elements are brought together to form a production facility it is easy to lose the efficiency each one may have on its own.

The Mixer

The mixer is the system of knobs, switches and sliders that is the heart of the facility. Products from various manufacturers may look alike at first glance. However, they can be quite different in detail. Look for good layout of most-used controls. On a larger mixer the operator may have to stand and lean to reach some of them. The switches and knobs that are used to configure the system for a project should be *presets*.

It is possible to have too many input channels. A simple rule suggests: *If you can't reach it, don't buy it.* Most mixer channels are on 1.5 in. centers. Thirty-six inputs are a reasonable limit to the average person's reach. Many input modules provide a selector to choose between two sources. A new generation of mini-mixers is on the market. These bring the controls closer together by using a monolithic rather than modular design. The hazard of this is that a single failure cannot be serviced without taking the whole package apart. However, these small mixers are a convenient and economical means of adding inputs when temporary expansion is required.

Another rule states: *If you can't see it, it isn't there.* Too many video production mixers become partially covered with scripts, schedules and other papers. The controls that are covered might just as well be gone. The provision of a sliding script holder is not a good choice. In a busy production room if the holder must be moved to change a setting, the operator may be half a page behind when the script is back in place. The best solution is a 12 in. apron between the operator and the mixer. It should be 1/2 in. below the lip of the mixer so that long pages will not easily slide up and push open some faders.

Height is important. If the equipment is too high, the operator is likely to raise the chair height. Unless a footrest is provided the user's upper legs will not be horizontal, leading to circulatory problems in the lower legs. The average desk or table is 30 in. high. A typewriter table is 27 in. high because the machine elevates the controls. The script apron is similar and provides full support for the operator's arms. Users not accustomed to this convenience soon learn its benefits over mixers with only a wrist rest.

Some mixers are built with a *scribble strip* between the faders and the input modules. Most faders have 1 in. of space at the top where labels can be placed. This makes the scribble strip area available for installation of controls that otherwise might be on the sides. If no scribble strip is provided, one or more input modules might be eliminated in favor of more convenient utility controls. Following is a list of functions that have been placed on these panels.

- Intercom
- Talkback
- Microphone on/off (privacy)
- Machine starts
- Announcer on/off
- On-air lights on/off
- Monitor stereo/mono
- Monitor dim and mute
- Phone coupler answer and release
- Silent signal lights
- Video monitor switching

Ancillary Equipment

In addition to the mixer there will be other equipment. Various players and recorders, processors and effects devices fill out the system for flexibility among projects. Some installations place these devices below and to the side of the mixer where the drawers of an office desk would be. This causes continual leaning and bending during normal operation, which is ergonomically unacceptable.

Ancillary equipment is better placed in a rack to the side of the operator from elbow height to just above eye

3.1 PLANNING AN AUDIO PRODUCTION CENTER

ROY W. RISING
ABC-TV, HOLLYWOOD, CA
ADDITIONAL MATERIAL PROVIDED BY MALCOLM M. BURLESON AND DAVID CARR
KHOU-TV, HOUSTON, TX

INTRODUCTION

Planning this chapter parallels planning an audio production center. One thing is likely—something will be overlooked. This chapter is a result of more than three decades of the author's involvement, both as a designer and an end-user of broadcast audio facilities. The subjects presented are those that should not be overlooked. Most of these will be covered in greater detail in following sections of this *Handbook.*

The objective of planning is to consider every possible detail and develop an integrated scheme that leads to the best final result. There will be tradeoffs; no single element is so important that it cannot yield to another's requirements. It helps to begin by giving equal weight to every aspect. In the end the larger and smaller issues will find their own places.

For example, every control room needs a pencil sharpener. Electric sharpeners are convenient but noisy and usually require a tabletop. Graphite dust is an enemy of electronic equipment because it conducts electricity. When the room is shut down, chances are the electric sharpener will not work until the correct circuit breaker is turned back on.

It may be better to choose a manual sharpener. It is quieter and can be wall-mounted away from the electronics. Experience has shown that installation near the entry of a room serves a third consideration. Personnel from adjacent areas can slip in and sharpen a few pencils without disturbing the operation. Those within the room should routinely sharpen two or more pencils each time. If the lead of a newly sharpened pencil breaks, find another brand.

Attention to this seemingly minor item reveals its larger importance. It is a good model for contemplation of the other details in systems design. An extension of the thinking mandates keeping a wastebasket below the sharpener.

Some things will be assumed. There should be an architect with a good understanding of acoustical considerations. A HVAC engineer will be needed to look after silent but effective ventilation systems. Keep in mind that although a solid-state device runs cooler than its vacuum tube ancestor, the equipment density is much higher and heat loads may be larger than expected. Electrical services are easily underestimated. The added cost of a larger entry capacity is small and will be welcome when future expansion presents new demands.

When initial sketches are completed and bids begin to arrive, establish a budget with about 20% slack for overruns. Watch expenditures for warning signs of underestimation. Setup a timetable based on realistic expectations. Track the progress and be alert for scheduling conflicts that might slow things down. Finally, if time permits and the budget is not depleted, have a party to toast your colleagues and yourself for a job well done.

USER INPUT

A very important part of the planning process is continuing dialog with the actual end-users. As plans for various elements of the facility begin to take shape, circulate sketches and descriptions among the users. Establish an open return channel for constructive criticism and follow-up for comments even when none are volunteered. Maintain a log of memos and conversations. This is helpful in resolving later disputes.

The end-users are justifiably nearsighted. Their focus is on those things that matter the most personally. Part of the planner's task is to moderate the areas where different user's needs overlap. First develop a consensus among users who share the same areas. Next design a proposed zone of transition between adjacent areas. Then circulate the proposal and make adjustments based on the responses.

LAYOUT

Very few production facilities have the good fortune to be built new from the ground up. More often an existing structure is used. By using an existing production center, broadcasters can benefit from the good aspects of the center and avoid its previous problems.

There are three types of audio production facilities: radio, TV and video postproduction. The audio control rooms and ancillary areas are similar and follow some common patterns. The first consideration for layout is sound isolation. Rooms with high level loudspeaker monitoring should not be adjacent to each other or to rooms with live microphones. The obvious exception is where a performance studio must be connected to

Section 3: Audio Production Facilities

Introduction by Randy Hoffner, ABC, New York, NY

Audio communication is the foundation of broadcasting. From early radio transmissions to 5.1-channel AC3 digital audio, sound has played a pivotal role in the growth of radio and, later, television. It was stereo that lifted FM from the "forgotten medium" of the 1960s to the prominent role that it enjoys today. Likewise, the DTV sound system—AC3—holds great promise for significantly enhancing the viewing experience of high-definition television. To enjoy the benefits of such advanced technologies, however, careful attention must be paid to the details of audio capture, storage, processing and transmission. The qualities and characteristics of aural signals have been hard to quantify, and even harder to preserve. Unlike video, where absolute references readily exist in the picture, audio is a tapestry of elements and nuances that will be irrevocably lost unless properly handled.

Roy W. Rising directs you through the planning of an audio production center from start to finish in his chapter, *Planning an Audio Production Center.*

Written by NAB Science & Technology Senior Vice President Lynn D. Claudy, the *Principles of Acoustics for Broadcast Applications* chapter offers a broad overview of acoustics, with a blend of theory and practice to provide readers with a balanced foundation on the subject.

Ty Ford's *Microphones* chapter is intended to provide engineers with information that will help them make the right choices. The chapter reviews the basic types of microphones, their construction and how their different characteristics can be used in various applications for best results.

As digital broadcast equipment is poised to eclipse analog studio systems, will your broadcast studio be ready for the change? Walt Lowery's *Studio Audio Equipment* chapter prepares broadcasters to take the leap into the digital future.

Michael Starling's *Audio Recording Systems* chapter illuminates the significant advances in audio compression techniques and the tremendous effects that such developments have had on the storage of digital audio.

David T. Turner examines the broad concept of station automation in his *Station Automation and Networking* chapter by providing a thorough overview of each component; thereby guiding the broadcast engineer through this seemingly daunting task.

The digital audio facility of the future will revolve around compressed signals. Therefore understanding the processing techniques is of great importance to broadcast engineers. Fred Wylie provides an overview of these technologies in his *Digital Audio Compression Technologies* chapter.

"One mark of the professionalism of a broadcast engineer is their mastery of the techniques of audio processing." Find out if you have what it takes in Robert Orban's *Transmission and Audio Processing* chapter.

Skip Pizzi and Jerry Whitaker's *Remote News and Production* chapter addresses the two most critical areas in remote broadcasting: portable production systems and backhaul.

Steve Church's *Telephone Network Interfacing* chapter explores the myriad of methods to integrate the telephone network into broadcast operations. Church covers the various services available from telephone companies and ways to interface them with station facilities.

Skip Pizzi contributes his expertise again in this section with his *Common Carrier Audio Program Services* chapter, which considers the opportunities for audio signal interconnections afforded to broadcasters by telephone companies and examines how broadcasters can implement these hookups.

Insulation Resistance Test

An insulation resistance test can be performed on a transmission line system after purging with dry gas to determine if it is sufficiently dry before applying power. The insulation resistance between the inner and outer conductors should be greater than 100,000 megohms. There are many inexpensive test units available to perform this test. This is especially critical when installing semi-flexible cable that has been stored outdoors for an extended length of time and it is unknown whether positive pressure has been maintained.

Operation and Maintenance

The following guidelines should be followed to ensure a long term and reliable operation of a coaxial transmission line system:

- Establish a baseline VSWR and time domain response of the system during installation
- Set and maintain all VSWR or reflected power protection devices at nominal levels. If a fault occurs, do not override these devices in order to continue transmitting at high power
- Always maintain positive pressure on the line system
- Do not exceed the pressure rating of the line or antenna. Damage to components and personnel may result
- Perform periodic maintenance checks on the pressurization system. Deviations from base line levels are a first indication of potential problems with the line
- Check for hot spots along the line during routine inspections. The outer conductor temperature is proportional to the inner temperature and excessive heating may indicate the beginning of a failure of the inner conductor connections
- Check the horizontal and vertical line sections for wear caused by thermal expansion
- Remeasure the system VSWR and time domain response on a regular basis and compare to previous results for changes.

Table 2.4-8
Recommended length of horizontal run.

Vertical Run (feet)	Recommended Length of Horizontal Run (feet) 3-1/8"	4-1/16"	6-1/8"	8-3/16"	9-3/16"
100	15	15	15	20	20
500	25	30	35	40	40
1000	35	40	50	60	60
1500	40	50	60	70	70
2000	45	60	70	80	80

all possible suppliers are aware of your system needs and individual component types so that the appropriate adapters can be supplied if needed.

MAINTENANCE AND TESTING

External air dielectric transmission lines must be kept under positive pressure using a dry gas to prevent moisture or other contaminants from entering the system. The moisture and contaminants can accumulate on the internal surfaces and connection joints of the transmission line causing oxidation and corrosion resulting in:

- increase in line attenuation
- increase in VSWR
- localized heating due to resistive losses
- voltage flashover at dielectric surfaces.

Transmission lines pressurized with dry gas maintain the initial performance of the system and reduce the risk of damage to the line and subsequent off-air situations. Dry gas is typically obtained from nitrogen bottles or a dehydrator system. Care must be taken not to exceed the lowest pressure rating of any of the system components.

Prior to applying power to the transmission line system, it should be sufficiently purged of environmental air acquired during installation. This normally can be accomplished by exchanging at least 3 volumes of dry gas through the system by opening a purge valve or joint at the far end of the system, and allowing the dry gas to pass through the system for a predetermined length of time based on the size of the system. Purging should also be performed anytime the system is opened to outside air or a positive pressure condition has not been maintained.

Electrical Testing

Electrical testing of the transmission line system should be performed for any of the following conditions:

- During initial installation
- After VSWR trips to determine the cause and location
- Periodically, to document performance
- At any time the performance of the system is in doubt.

Most conditions that result in less than optimal performance of the transmission line system can be detected immediately after installation. The most useful tests to determine performance are the VSWR/return loss sweep test and RF pulse testing. There are many types of equipment that can perform these tests. Over the past few years, the network analyzer with time domain transform capabilities has become the standard type of equipment.

The VSWR/return loss test is used to determine the composite input VSWR of the transmission line system terminated by a load or antenna as seen by the output of the transmitter. Normally, measurement is performed of the VSWR as a function of the frequency within the broadcast channel of interest; however, wider frequency sweeps are recommended to determine possible performance over wide temperature variations due to line length changes from thermal expansion/contraction or antenna VSWR changes due to environmental conditions such as icing.

Since the VSWR test is a composite, it presents an overall view of the system performance. System specifications for VSWR must take into account all the individual components, therefore, long complicated systems will typically have higher VSWR responses than short simple systems. Improvements can be made through the use of fine tuners installed near known contributors of VSWR, such as elbow complexes. Exceptional care must be taken during installation and tuning at the factory of elbows, adapters and transformers that will be used.

The RF pulse test displays reflections in the channel response as a function of time rather than frequency. This can be of great help in locating the cause of excessive VSWR or faults within the system. It is also used to identify the reflections from the antenna and other far end components to determine if ghosting conditions may be present. A typical far end reflection criteria to prevent most problems with the transmitted signal is −33 db return loss within the broadcast channel of interest.

Where the cause of high VSWR may be an improperly assembled flange or damaged conductor, the in channel RF pulse does not have the resolution to accurately locate the fault. Since the width of the pulse is inversely proportional to the frequency bandwidth, a narrower pulse for resolution requires wider bandwidth. With network analyzers presently on the market, the feature to change the bandwidth is readily available. These can be used in similar fashion to the time domain reflectometry (TDR) test. Bandwidths of 50–150 MHz centered on the channel will usually provide sufficient resolution to identify the above types of faults. It should be remembered that this test is not to determine the in channel characteristics of the system but rather to provide a more detailed look at the transmission line. In particular, components that are frequency sensitive, such as an antenna, may present a significant reflection when using wider bandwidths. These types of components should be identified ahead of time to prevent incorrect analysis of the data.

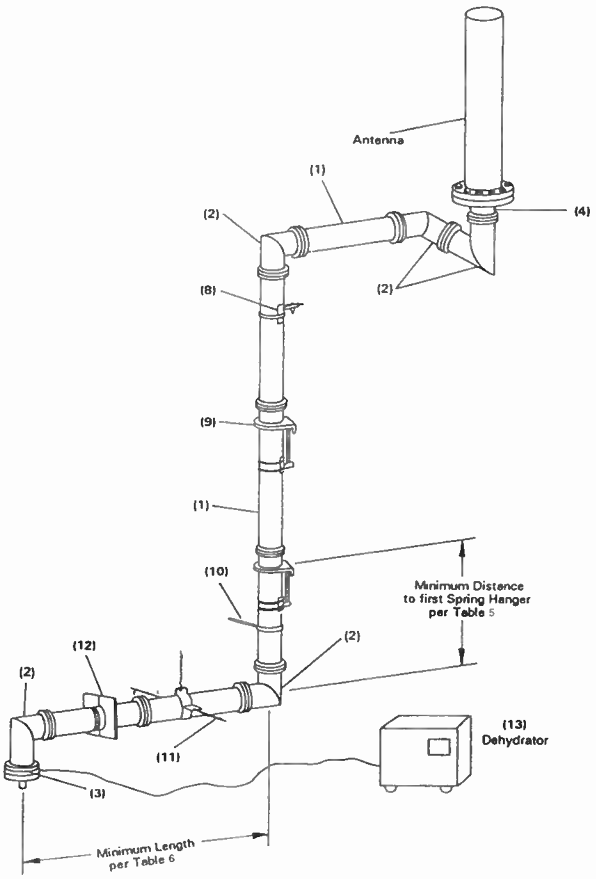


Figure 2.4-12. Typical broadcast rigid line system.

point of entry to the building, horizontal movement in and out of the tower must also be accounted for during installation (see Table 2.4-7 and 2.4-8).

Line Interfaces

Interface standards have been established by the Electronics Industries Association (EIA) for 50 Ω lines (RS-225) and 75 Ω lines (RS-259). The user should be aware, however, that many proprietary designs are used by suppliers that are not completely compatible with the EIA standards or other manufacturer's components. When purchasing components from different suppliers or making repairs to a system, be sure that

Table 2.4-7
Minimum distance to first vertical spring hanger versus length of horizontal run (refer to Fig. 2.4-12).

Horizontal Run (feet)	Minimum Distance to Lowest Hanger (feet) 3-1/8"	4-1/16"	6-1/8"	8-3/16"	9-3/16"
20	5	6	9	12	13
40	6	7	11	15	17
60	7	8	13	17	20
80	8	9	14	19	22
100	9	10	16	21	23

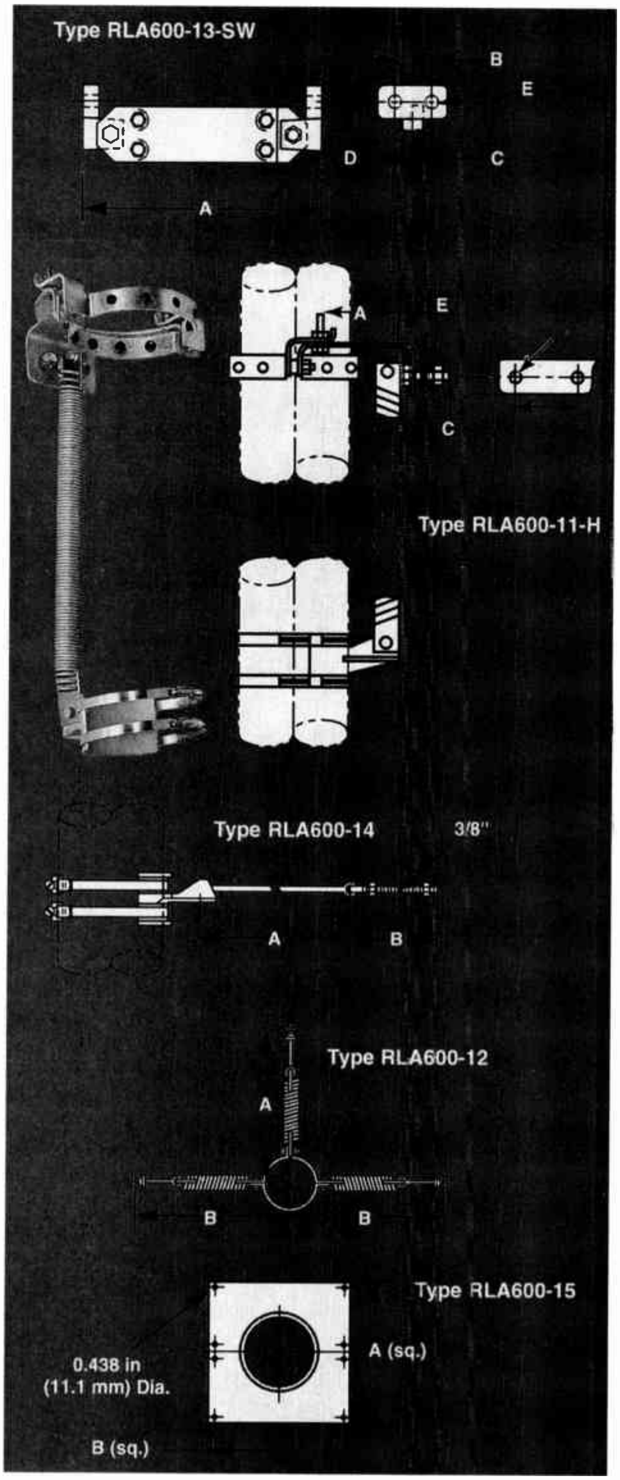


Figure 2.4-13. Detail of spring hanger elements.

the movement to fall inside the inner conductor and not out into the main portion of the line. The spring is designed to be self-compensating in this process providing a constant contact pressure on the inner conductor. However, the mechanical wear does limit the reliable life of the spring, based on the conditions of operation for that particular line system.

The bellows device is built directly into the inner conductor (see Figure 2.4-10). Thermal compensation is achieved by the expansion and contraction of the corrugated section similar to the principal used in the construction of semi-flexible lines. The benefit of this design is that the bellows absorb all of the inner conductor movement and eliminate any wear at the bullet/inner conductor interface. The reliable life of the line is determined by the number of expansion/contraction cycles the bellows is designed to withstand.

The continuously corrugated inner conductor concept works on exactly the same principal as semi-flexible line. The entire inner conductor is corrugated allowing significant expansion and contraction capabilities (as shown in Figure 2.4-11). The inner conductors can now be bolted together with no possibility of a sliding contact point. Because the bolting process produces higher contact pressures, a significant reduction in resistance can be obtained resulting in a better electrical connection. In effect, after assembly of this type of system, there is one continuous inner conductor rather than multiple inners that have been spliced together.

The different compensation methods must be evaluated by the user relative to the specific environmental and operating conditions present at the broadcast site. Factors such as installation, maintenance requirements, extreme ambient temperatures, and high operating power levels can be significant in the determination of the proper concept needed for long term reliability at minimum costs.

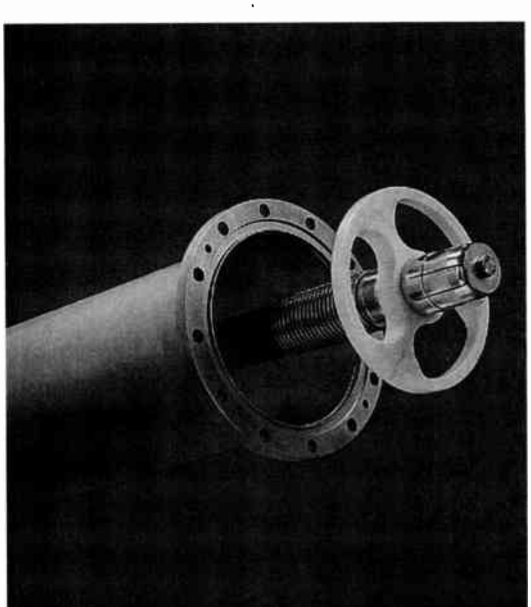


Figure 2.4-10. Bellows compensation approach to line length changes.

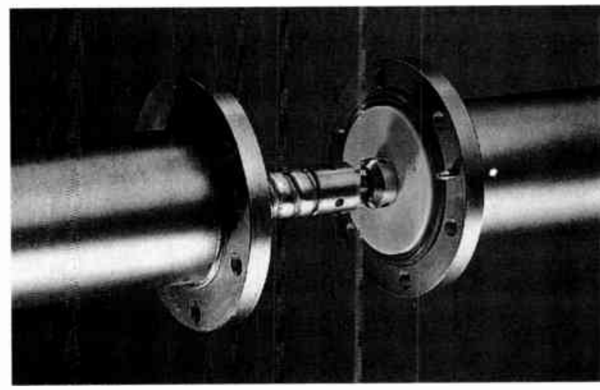


Figure 2.4-11. The use of a corrugated inner conductor for rigid line.

Rigid Line Systems

A rigid coaxial transmission line system is composed of a vertical exterior run and a horizontal run that will have some components inside the transmitter building and some attached to an exterior horizontal support bridge between the building and the tower. Because rigid lines are air-dielectric types, all components subject to contamination by outside air must be under constant pressurization.

Differential Thermal Expansion

The most important factor to consider in a rigid line installation is the difference in thermal expansion coefficients between the copper line and the steel support tower. For a 100°F change in ambient temperature, the differential expansion between the line and tower will be about 0.4 in. per 100 ft of length. For a 1000 ft. tower and a temperature change from 0°F to 100°F, this will result in the rigid line growing 4 in. relative to the tower. Provisions must be made to support the line within the tower and at the same time allow it to float to prevent severe buckling strains on the tower and line. This is accomplished through the use of vertical spring hangers which attach rigidly to the tower but support the weight of the line by springs and clamps (as illustrated in Figures 2.4-12 and 2.4-13).

To account for ice loads and possible failures of the spring hangers, a rigid hanger is attached near the top of the vertical run. It is important that it be placed as close to the antenna input as possible to prevent any thermal expansion conditions causing excessive force on the antenna input. Since the rigid connection is at the top of the run, the cumulative effects of the thermal expansion or contraction occur at the bottom of the vertical run. This is where care must be taken to allow sufficient clearance for the line to enter the tower from the horizontal run and not be damaged by tower members as the line moves up and down relative to the tower. And since the horizontal run is anchored at the

Table 2.4-5
EIA flange hardware requirements and recommended torque values

EIA Flange Size	Bolt Size	No. of Bolts	Recommended Torque Value
7/8"	1/4"	3	80 lb-in (9.0 N-m)
1-5/8"	5/16"	4	140 lb-in (15.8 N-m)
3-1/8"	3/8"	6	20 lb-ft (27.1 N-m)
6-1/8"	3/8"	12	20 lb-ft (27.1 N-m)
8-3/16"	3/8"	18	20 lb-ft (27.1 N-m)
9"	3/8"	20	25 lb-ft (34.0 N-m)
9-3/16"	3/8"	20	20 lb-ft (27.1 N-m)

produces a small reflection. At some frequency, the distance between flanges will become periodic and a VSWR spike will be generated. This spike will typically have a narrow frequency response but can be significant based on the number of discontinuities and the magnitude of the reflections. The frequency of the VSWR spike is determined by the following equation:

$$F_{\text{spike}}\,(\text{MHz}) = \frac{492.15 * V_p * N}{L}$$

where:

F_{spike} = Frequency of spikes (MHz)
N = Any integer
V_p = relative velocity of propagation
L = equal distance between discontinuities (ft.)

The response curve for this equation has a bandwidth of approximately ±2 MHz. Therefore, the line type must be chosen so that a critical frequency of operation is a minimum of 2 MHz from the spike frequency. Since many rigid line systems are several hundred feet in length, the amplitude of the VSWR spike can be quite large. Due to this, some manufacturers allow greater safety margins from the critical frequency when determining rigid line section lengths.

This equation is used to determine standard line lengths for rigid lines as shown in Table 2.4-6.

Internal Thermal Expansion

In a high power broadcast application, the inner conductor will run substantially hotter than the outer conductor. A typical difference in operating temperatures between the inner and outer is 70°F.

Because of this temperature difference, the inner conductor will expand to a longer length relative to the outer conductor. This difference can be calculated based on the thermal expansion of copper being 9.8×10^{-6} in/in/°F. With an overall length of 20 ft, the difference in lengths will be 0.174 in. In semi-flexible line, this difference would be absorbed by the corrugations in the conductors. For rigid lines, however, it must be compensated for at the inner conductor and inner connector (bullet) interface.

Without the thermal expansion, the inner conductor would be cut to exactly mate up to a shoulder on the bullet. This provides the optimum electrical contact and impedance match. To account for thermal expansion, the inner is cut back the distance needed to allow for the expansion under the maximum allowable operating temperature differential between the outer and inner. So, as the temperature varies between input power levels and ambient temperatures, the inner will move back and forth on the bullet. This will cause material deterioration from friction and ultimately will result in either mechanical or electrical failure at the contact points.

To accommodate this movement but minimize the degradation of the contact surface, compensation devices have been developed for rigid lines. They come in three concepts:

- Watchband spring attachment to the bullet
- A bellows attachment to the inner conductor
- A completely corrugated inner conductor using bolted joints between each section.

The watchband spring concept provides for a spring-like contact device that is attached circumferentially around one side of the bullet (as shown in Figure 2.4-9). The spring side is then inserted at the top of the lower inner conductor (the line section closer to the transmitter) such that the spring is below the insertion point. The insertion force of the upper side of the bullet is such that during thermal changes, the inner conductor movement is primarily on the lower spring section. This allows any shavings that are caused by

Table 2.4-6
Recommended rigid line section lengths

20′ Section	19′ 9″ Section	19′ 6″ Section	19′ Section
TV Channels			
2, 3, 5, 6, 7, 8, 11, 12, 14, 15, 18, 19, 23, 27, 31, 35, 39, 40, 43, 44, 47, 48, 52, 56, 60, 64, 68	16, 20, 24, 28, 32, 33, 36, 37, 41, 45, 49, 53, 57, 58, 61, 62, 65, 66, 69	4, 9, 10, 13, 17, 21, 22, 25, 26, 29, 30, 34, 38, 42, 46, 50, 51, 54, 55, 59, 63, 67	
FM Radio			
88.1–95.9 MHz 100.3–107.9 MHz		96.1–98.3 MHz	98.5–100.1 MHz

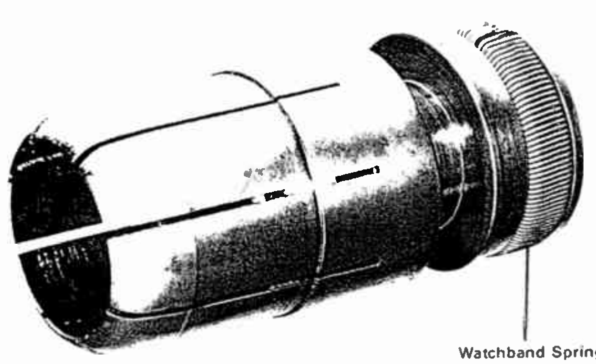


Figure 2.4-9. The watchband spring approach to thermal expansion.

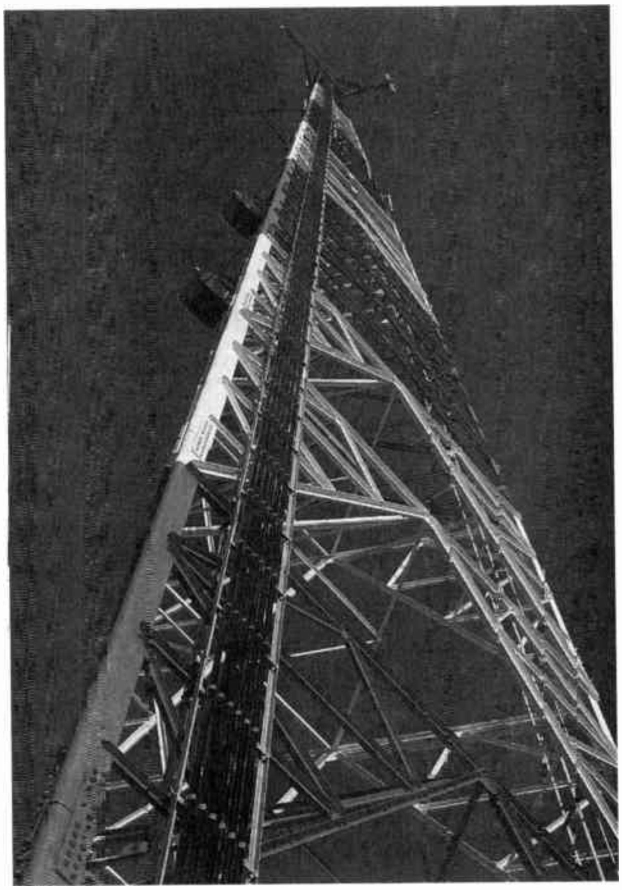


Figure 2.4-8. Mounting of coaxial cable lines on the face of a tower.

stranded wire is not recommended as they tend to deteriorate with time, reducing the effectiveness of the grounding connection. After installation, all ground connections should be weatherproofed to prevent degradation.

Phase Stabilized Lines

In some cases, such as AM directional antennas, it is necessary to use coaxial cables to sample current levels over time to confirm compliance with a broadcasting license. In this case, the phase stability of the cable over time is extremely important. Ordinary coaxial cable should not be used in these instances since it is subject to significant phase variation with temperature change during its initial usage. Instead, phase stabilized foam polyethylene dielectric or air dielectric cables should be specified. This cable has been subjected to a process that cycles the cable through a wide temperature range to remove the phase instability caused by hysteresis.

Rigid Line

Rigid lines have inherently high power handling and low attenuation and VSWR which make them ideal for high power broadcast applications. The line is typically fabricated from high conductivity oxygen-free hard copper tubing in nominal lengths of up to 20 ft. The inner conductor is supported in the outer conductor by peg or disk insulators. A PTFE compound is normally used for the insulating materials due to its extremely good electrical and mechanical properties.

The inner conductors of adjacent line sections are joined together by inner connector "bullets." These act as splices between the inner conductors and are designed with tension spring fingers to provide high insertion forces that produce low resistance electrical contacts. The high insertion force is also needed for good thermal conductivity. For added conductivity, the bullets are normally silver plated. An insulator is also attached to the middle of the bullet in order to anchor it between the outer conductor flanges. This provides mechanical stability when installing the line in a vertical position.

As mentioned above, the outer conductors are normally attached by bolting flanges together. The RF contact is accomplished through a raised contact surface near the inside diameter of the flange. A pressure seal is obtained through the use of an o-ring between the flanges. Flange types are typically governed by EIA standards RS-225 and RS-259, however, proprietary designs of various suppliers are also available. Important parameters for rigid coax are given in Tables 2.4-4 and 2.4-5.

Rigid Line Section Lengths

Because the connections between rigid line sections present an imperfect impedance transition, each flange

Table 2.4-4
Characteristics for commonly used rigid lines

Nom. OD of Outer Condr.	Zc (ohms)	Maximum Freq. (MHz)	Vel. of Prop (Percent)	Nom. OD of Inner (Inches)	Nom ID of Outer (Inches)	Net Weight (lb/ft)
7/8"	50	6000	99.8	.341	.785	0.6
1-5/8"	50	3000	99.8	.664	1.527	1.3
3-1/8"	50	1588	99.8	1.315	3.027	3.0
4-1/16"	50	1197	99.8	1.711	3.935	5.6
6-1/8"	50	788	99.8	2.600	5.981	7.3
6-1/8"	75	900	99.7	1.711	5.981	6.75
8-3/16"	75	709	99.7	2.293	8.000	9.0
9-3/16"	50	530	99.7	3.910	9.000	11.45
9-3/16"	75	600	99.7	2.580	9.000	11.45

warpage. Application of silicone grease to the o-ring to hold it in place is a common practice. However, it should be remembered that silicone is an insulating material and excessive application of the grease can result in contamination of the contact surfaces. Excessive heating and damage to the o-ring can result.

Installation

A proper installation can be the difference between a smooth running transmitter site and a continuous effort to correct failures in the transmission line and antenna system. It should be remembered that the environmental forces acting on transmission lines installed outdoors and on tall towers are constantly changing. This results in extreme stress on the cable from corrosion, vibration, atmospheric contaminants and thermal changes that are not present with indoor installations. Following the manufacturer's recommended procedures for installation and maintenance can prevent catastrophic failures.

General guidelines for installation of semi-flexible cable systems primarily center around the handling of the cable. Upon delivery, the cable is usually shipped on reels or in cartons which should be inspected for shipping damage. In the case of air-dielectric cables, check to see that the cables are still holding pressure. In storage, allowable pressure drop is typically 1 psig (70 kPa) per 24 hours. If there is excessive leakage, the cause should be determined prior to installation. Reels should always be stored and moved on the flange edges. Storing on a flange side or dropping a reel can cause damage to the inside lengths of cable due to crushing from the outside.

The actual lifting and installation of the cable should only be performed by qualified installation crews. The proper use of lifting devices, such as hoisting grips, winches, load lines, etc. is required for safety (see Figure 2.4-7).

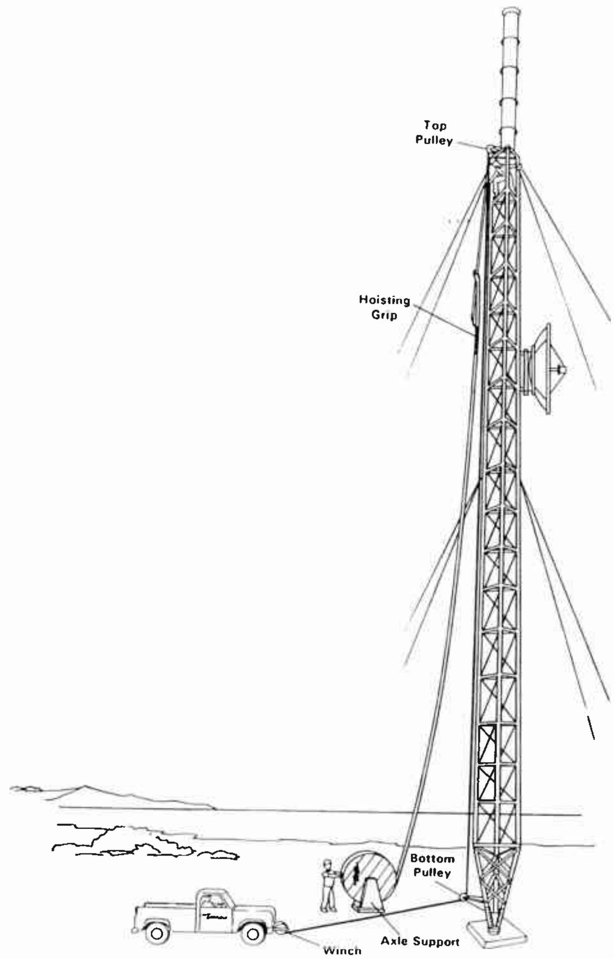


Figure 2.4-7. Installation of coaxial cable.

For long term reliability, proper installation of cable hangers is essential. The purpose of the hangers is to support the cable's weight and to prevent movement from wind, rain, ice and other forces. Excessive movement causes abrasion and damage to the outer conductor. To begin, the manufacturer's recommended hanger spacing should be considered. This spacing should be modified (typically twice as many hangers used) if the wind velocity will be consistently above 40 mph, peak winds exceed 125 mph or heavy ice is expected to form. (A common support system is shown in Figure 2.4-8).

A drip loop is recommended in the horizontal run to prevent moisture from using the line as a path to the building wall. Water can travel down the outside of the jacket and build up at the wall feed-through if a drip loop is not used and the line slopes toward the building. Water may also travel between the jacket and outer conductor if damage to the jacket has occurred. If this occurs, a small notch cut into the jacket at the bottom of the drip loop will allow for drainage.

Because the jacketed copper cables are very corrosion resistant, it is also possible to bury the horizontal run either directly or pulled through a conduit. Sharp objects must not come into contact with the cable or damage to the cable may occur. If installing within a conduit, do not exceed the tensile strength of the cable when pulling it through. Conduit pulling lubricants should be applied generously to the cable. Bends should be avoided if at all possible as the corrugations may cause excessive friction when pulling around a corner. Since a buried cable may occasionally become submerged, any buried connectors or splices should be well protected with a good weatherproofing kit.

Grounding

Since the cable represents a direct connection between the outside and the transmitter equipment, proper grounding is necessary to protect the equipment from the large currents caused by a lightning strike. Minimum grounding conditions are to ground the cable at the top and bottom of the vertical run as well as just before entry to the building. For long runs, additional grounding in 200 ft increments is recommended.

Ground connections must be made with high quality copper wire or straps. Braided copper straps or fine

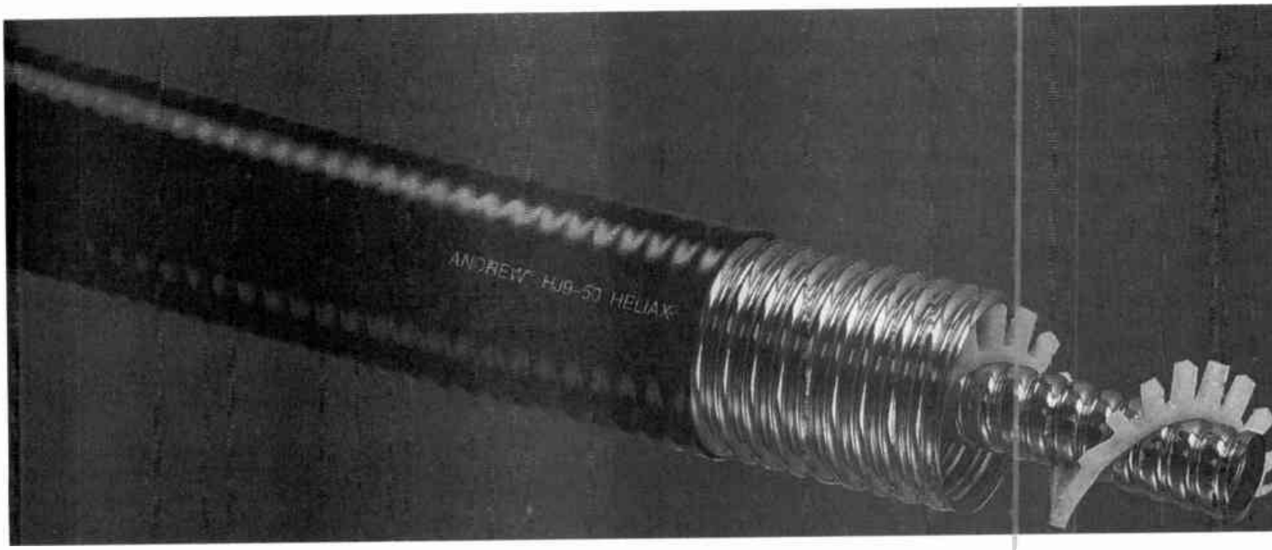


Figure 2.4-6. A section of high-power air dielectric semi-flexible coax. *(Courtesy of Andrew Corporation.)*

Connectors, Splices and Adaptors

Usually, the connector type is determined by the antenna and transmitter connection type and the power handling requirement of the transmission system. Once the decision has been made, then it is important to make the attachment properly. Close attention to the supplied assembly instructions, in particular the cable trimming dimensions, will assure a good connection. Clean mating surfaces and proper tightening is required for good electrical contact along the entire 360° of the cable edge.

It is highly recommended to make all connector and splice attachments with the cable end in a horizontal position. This minimizes the chance of metal chips falling into the cable during the trimming process. This will normally mean making connections on the ground, however, there may be times when there is no choice but to make the attachment on the tower. In that case, it is important to take precautions, such as stuffing cloth into the open cable end to prevent chips from falling into the cable. Care must be taken when removing the cloth so as not to drop any chips into the now open cable end.

Proper attachment of the finished connector to the antenna or transmitter prevents loosening of the connection from vibration or other stresses. It is important to recheck the cable assembly after attachment, specifically for connections using screw together interfaces. Twisting of the cable during attachment may cause the connection to loosen and result in poor performance and environmental damage to the cable. These connections should be securely taped to prevent loosening.

EIA style flange connectors must be carefully tightened to prevent distortion of the flanges. Tighten bolts that are opposite (not adjacent) to the proper torque specification. Overtightening can also result in flange

Table 2.4-3
Characteristics of typical coaxial cables

Cable	Size	Max. Freq. (GHz)	Velocity (Percent)	Nominal Inside Transverse Dimension (cm)	Dia/over Jacket (Inches)	Min. Bend Radius (Inches)	Weight (lb/ft)
Foam-Dielectric Cable							
	7/8"	5.00	89.0	2.11	1.09	10	0.33
	1-1/4"	3.30	89.0	3.11	1.55	15	0.66
	1-5/8"	2.50	88.0	4.05	1.98	20	0.92
Air-Dielectric Cable							
	7/8"	5.20	91.6	2.02	1.11	10	0.54
	1-5/8"	2.70	92.1	3.99	1.98	20	1.04
	1-5/8"	2.70	92.1	3.99	1.98	20	1.04
	2-1/4"	2.30	93.1	4.96	2.38	22	1.16
	3"	1.64	93.3	6.35	3.02	30	1.78
	3-1/2d"	1.43	96.0	7.52	3.50	30	1.98
	4"	1.22	92.0	8.55	4.00	40	2.50
	5"	0.96	93.1	11.30	5.20	50	3.30
	6-1/8"	0.86	97.0	14.70	6.73	79	7.33
	9"	0.65	97.0	19.50	8.90	98	12.50

Table 2.4-2
Common types of solid outer conductor cables.

Cable Designation	Nominal Z_0	Vp	Diameter (in)	Nominal Atten. @ 50 MHz Db/100 ft	Nominal Atten. @200 MHz dB/100 ft	Nominal Atten. @700 MHz dB/100 ft	Max. Op Voltage (RMS)
FSJ2-50	50	.84	.290	1.27	2.58	4.97	6400
FSJ2-50	50	.83	.415	.848	1.73	3.37	13200
FSJ4-50B	50	.81	.520	.73	1.50	2.97	15600
LDF2-50	50	.88	.44	.736	1.50	2.93	15600
LDF4-50A	50	.88	.63	.479	.983	1.92	40000
ETS1-50T	50	.82	.29	1.27	2.56	4.89	6400
ETS2-50T	50	.83	.415	.856	1.77	3.48	13200
FSJ1-75	75	.78	.29	1.3	2.68	5.30	3300
FSJ4-75A	75	.81	.52	.673	1.39	2.75	10000
LDF4-75A	75	.88	.63	.435	.896	1.76	26000

systems. The line achieves its flexibility through the use of corrugated copper conductors and can be fabricated in extremely long lengths. This eliminates the need for interconnecting joints therefore minimizing possible installation/assembly problems. They are also easier to install because the corrugations eliminate most of the need for hanger designs that permit independent movement of the line from the tower (differential thermal expansion).

Foam-Dielectric Cable

Foam-dielectric cables are designed for systems that do not require a pressure path to the antenna or other connected components. Typical applications include radio and low power television. Sizes range from 1/4" to 2-1/4". These cables are constructed with a closed-cell, low density foam dielectric which prevents water penetration while providing low attenuation and high relative velocity of propagation (see Figure 2.4-5).

The maximum diameter of foam-dielectric cables is limited by the manufacturing process of "coating" the inner conductor with the foam dielectric. The attachment process requires the foam to be applied to the inner in a liquid state and immediately cooling the foam to prevent distortion of the foam shape and movement of the inner conductor within the foam. As the diameter of the foam and inner conductor increases, the outer portion of the foam will harden and act as a thermal insulator for the inner portion of the foam. This action is much like the theory behind styrofoam insulated coolers. Whatever is inside the cooler is insulated from outside forces. Now, because the inner portion of foam is not hardened, the actual weight of the inner conductor can cause the conductor to sag in the liquid foam, resulting in a non-concentric configuration of the conductors. The characteristic impedance has now changed and the electrical performance has been compromised.

To produce larger diameter cables requires the use of spiral sections of dielectric wrapped around the inner conductor. This results in what is called an air dielectric cable.

Air-Dielectric Cable

Air-dielectric cables utilize a spiral dielectric material to separate the conductors. Because the majority of the volume between conductors is now air, these cables have slightly lower attenuation and higher average power handling than foam-dielectric cables. They require pressurization not only to prevent water ingress but also provide pressure to the antenna if needed. A benefit of the pressurization is that significant damage to the cable will usually result in a pressure leak allowing the damage to be detected before a more severe condition occurs as a result of voltage breakdown.

A spiral-wrapped air dielectric cable is shown in Figure 2.4-6. Table 2.4-3 compares common cables used in broadcasting.

Figure 2.4-5. A common type of foam-filled semi-flexible coaxial cable using a hollow center conductor. *(Courtesy of Andrew Corporation.)*

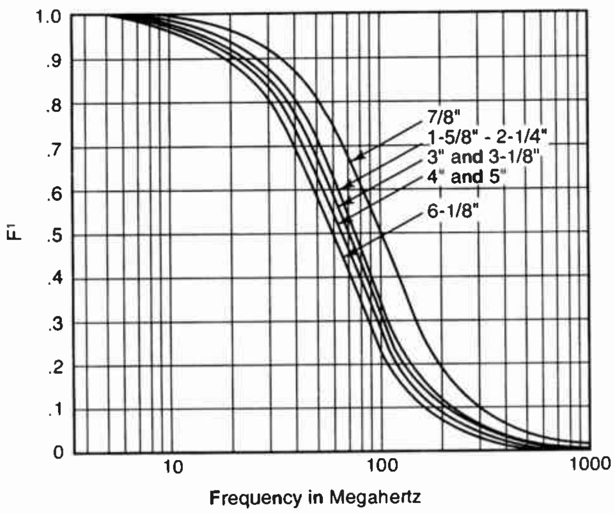


Figure 2.4-3. Derating factor for average power due to VSWR.

tween two or more lines is the velocity of propagation or phase velocity (V_p). It is expressed as a fraction of the speed of light in a vacuum and is determined by the dielectric constant of the insulating material:

$$V_p = \frac{c}{\sqrt{\epsilon'}}$$

$$\frac{1}{\sqrt{\epsilon'}} = \text{relative phase velocity}$$

where c = speed of light

The result of adding a dielectric material other than air is to "slow down" the TEM wave. For two coaxial lines having a different percentage of dielectric material there will be a difference in electrical phase between them even though they are physically the same length. This can be very important when using cables for sampling the relative phases of an antenna array, for example. Most manufacturers supply this factor with the other specifications for a transmission line.

Figure 2.4-4. A common type of semi-flexible foam-filled coaxial cable. *(Courtesy of Andrew Corporation.)*

Flexible RF Cables

Coaxial transmission lines used for short interconnections between equipment, typically inside buildings, have traditionally been called RF cables. They are small in diameter making them very flexible and very useful in areas with minimum space. The inner conductor is usually solid copper or a copper clad metal and the outer conductor a copper or aluminum braid. In order to maintain concentricity of the inner and outer conductors, the dielectric is a solid insulating material. The more commonly used cables are shown in Table 2.4-1.

With the introduction of more digital signals and multiple frequency grouping, the shielding performance of braided outer conductors may not be sufficient to prevent signal leakage. Interference between cables within an installation can be detrimental to signal quality and a frustrating condition to track down the cause. In these critical installations, cables that utilize a solid outer conductor will improve the system performance. Typical cables are listed in Table 2.4-2 and shown in Figure 2.4-4.

Semi-Flexible Cable

Globally, semi-flexible cables are the most popular choice for broadcasting applications. It is ideal for a wide variety of low and medium power transmission

Table 2.4-1
Common types of flexible RF cables

Cable Designation	Nominal Z_0	Vp	Diameter (in)	Nominal Atten. @ 50 MHz Db/100 ft	Nominal Atten. @200 MHz dB/100 ft	Nominal Atten. @700 MHz dB/100 ft	Max. Op Voltage (RMS)
RG-8A/U	52	0.66	0.405	1.6	3.2	6.5	4000
RG-8/X	50	0.78	0.242	2.5	5.4	11.1	600
RG-213/U	50	0.66	.405	1.6	3.2	6.5	5000
RG-58/U	53.5	0.66	.195	3.1	6.8	14.0	1900
RG-58A/U	50	0.66	.195	3.3	7.3	17.0	1900
RG-58C/U	50	0.66	.195	3.3	7.3	17.0	1900
RG-11A/U	75	0.66	.405	1.3	2.9	5.8	5000
RG-59B/U	75	0.66	.242	2.4	4.9	9.3	2300
RG-62B/U	93	0.84	.242	2.0	4.2	8.6	750
RG-71/U	93	0.84	.245	1.9	3.8	7.3	750
RG-141A/U	50	0.695	.190	2.7	5.6	11.0	1400
RG-178B/U	50	0.695	.070	10.5	19.0	37.0	1000
RG6A/U	75	0.66	0.332	1.9	4.1	8.1	2700

Average Power

The average power rating is limited by the amount of heat created due to line losses. The amount of heat allowable is primarily determined by the safe, long term performance of the dielectric material used. Since the loss and temperature rise of the inner conductor are greater than the outer conductor, the maximum allowable temperature of the inner conductor is normally used to determine the average power rating.

For typical conditions, the inner conductor is allowed to reach a temperature of 100°C with an ambient temperature of 40°C. This means the inner conductor temperature is allowed to rise 60°C above the ambient. Based on this standard condition, the average power can be calculated from the following:

$$P_{avg} = \frac{16{,}380 * \sigma * D}{M_{\alpha} * \alpha} \text{ watts}$$

where

P_{avg} = average power rating for 60°C rise of inner conductor temperature
D = outer conductor OD, in.
σ = heat emissivity coefficient of outer conductor, watts/in^2
M = correction factor for attenuation (relative to 20°C)
α = attenuation constant, db/100 ft at 20°C

It should be noted that the heat emissivity coefficient is derived from experimental data and there are no industry standards as to its value for various line sizes. Therefore, it is possible to have noticeable differences between published ratings due to slight differences in this factor. For example, a difference of only one hundredth of a watt/in^2 can result in a 10% difference in the average power rating.

Typical heat emissivity values for 50 ohm rigid transmission lines:

7/8"	0.120 watts/in^2
3-1/8"	0.107 watts/in^2
4-1/16"	0.104 watts/in^2
6-1/8"	0.097 watts/in^2

The average power is also frequency sensitive since the attenuation constant (α) is frequency dependent. At higher frequencies, the attenuation is greater, therefore the average power rating will be reduced.

If average power is the most significant factor in choosing a transmission line, it is suggested to review data sheets from several manufacturers and discuss any differences in average power ratings. Other factors that should be reviewed are flange connection types, inner connector design and method of thermal compensation for the inner conductor. In this way, all factors that may affect the long term performance of the transmission line are addressed and a sound engineering decision can be made.

For broadcasting, the average power is dependent on the nominal CW power of the transmitter and the type of modulation. For the primary broadcast modulation schemes, the average power can be calculated from:

$$\text{AM} \quad P_{avg} = P_c\left(1 + \frac{M^2}{2}\right)$$

$$\text{FM} \quad P_{avg} = P_T$$

$$\text{TV} \quad P_{avg} = (0.6 + AU)\,P_{TV}$$

P_{avg} = average transmitter power
P_c = carrier power
M = amplitude modulation index
P_T = FM transmitter power
P_{TV} = TV peak sync power
AU = Aural to visual ratio

As with peak power, the average power rating must also be adjusted for actual operating conditions. Since average power is based on ambient temperature, it must be adjusted for the actual operating ambient temperature. (Use Figure 2.4-2 to find the appropriate adjustment factor.)

Average power should also be derated for VSWR. The derating factor (*DF*) is calculated from the formula below where F^1 is a factor that varies with frequency and line size. Select the F^1 factor from Figure 2.4-3, calculate the *DF* and then divide the average power rating by the *DF*.

$$DF = \frac{\text{VSWR}^2 + 1}{2 * \text{VSWR}} + \frac{F^1 * (\text{VSWR}^2 - 1)}{2 * \text{VSWR}}$$

Increases in average power ratings can also be theoretically accomplished in a similar fashion as peak powers. The same cautions apply for average power as did for peak power.

Velocity of Propagation

A final performance characteristic that has primary importance when attempting to match the phases be-

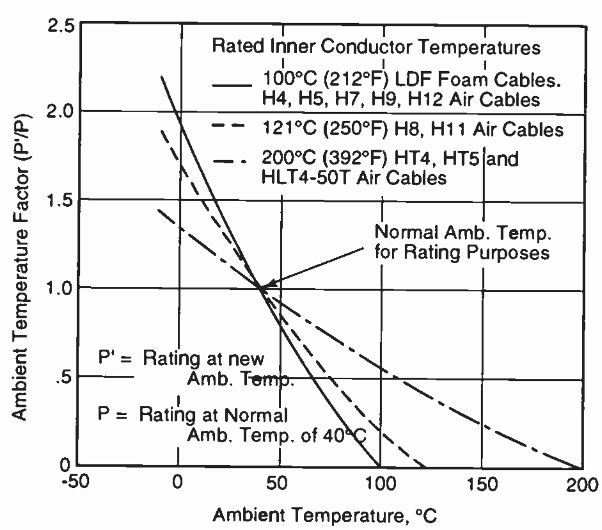


Figure 2.4-2. Variation of average power rating with ambient temperature.

modulation cycle. "Peak" refers to the peak amplitude of modulation and is not the instantaneous power when the RF voltage is at a maximum. In a continuous CW carrier such as FM, the peak power equals the average power. In 100% AM, the power rises to four times the carrier power at the peaks of the modulation envelope. So in this case, the peak power is four times the carrier power.

The peak power rating is dependent on voltage breakdown considerations, which are relatively frequency insensitive. Therefore, this rating is constant with frequency. It is, however, extremely sensitive to line size, physical conditions, pressure and the dielectric medium. It is important that the rating be based on a predetermined set of conditions that are clearly stated. Once a rating is determined for these conditions, it can be adjusted for other conditions.

The procedure for determining a peak power rating is to establish a peak voltage the line will withstand every time under normal manufacturing processes. Because voltage breakdown levels are extremely sensitive to effects such as dust, insulator condition and surface irregularities, the theoretical breakdown gradient cannot be used in practice. It has become standard procedure to derate the theoretical breakdown to 35% of its value to determine a production dc test voltage. This test voltage (E_p) is calculated using the following equation which is derived from the maximum voltage gradient in a coaxial line and includes the derating of 35%:

$$E_p = 3.17 * 10^4(d\delta)\left[\log\left(\frac{D}{d}\right)\right] * \left(1 + \frac{0.273}{\sqrt{d\delta}}\right)$$

where

E_p = production test voltage
d = inner conductor OD, in.
D = outer conductor ID, in.
δ = air density factor = 3.92B/T
B = absolute pressure, cm of mercury
T = temperature, °K

($\delta = 1$ for $B = 76$ cm and $T = 23°C = 296°K$)

Rounded off values for E_p for 50 Ω transmission lines are:

E_p DC Volts	Nominal Line OD inches
6,000	7/8
11,000	1-5/8
13,000	2-1/4
16,000	3
19,000	3-1/8
21,000	4
27,500	5
35,000	6-1/8

The production test voltage must now be converted to realistic RF rms operating voltage, E_{rf}:

$$E_{rf} = 0.7 * E_p * \frac{1}{\sqrt{2} * SF}$$

E_{rf} = maximum RF rms operating voltage with no derating for VSWR or modulation, but includes a safety factor

$1/\sqrt{2}$ = RMS Factor
0.7 = DC to RF factor
SF = safety factor for voltage (typically 1.4 or 2)

This voltage then determines the peak power rating, P_{pk}:

$$P_{pk} = (E_{rf})^2/Z_o \text{ watts}$$

or

$$P_{pk} = \frac{\left(\frac{E_p * 0.707 * 0.7}{SF}\right)^2}{Z_o}$$

An adequate safety factor on peak power is necessary to safeguard against voltage breakdown which can result in permanent damage to the line. Many manufacturers have settled on a power safety factor of 2 which is equivalent to a voltage safety factor (SF) of 1.4. When good conditions for installation or operation can not be maintained, it is good practice to apply a higher safety factor to prevent damage on initial startup of a system.

Once the peak power is determined, it must be further derated for the effects of modulation and VSWR. These deratings are calculated as follows:

AM:

$$P_{max} < \frac{P_{pk}}{(1 + M)^2 * \text{VSWR}}$$

FM:

$$P_{max} < \frac{P_{pk}}{\text{VSWR}}$$

TV:

$$P_{max} < \frac{P_{pk}}{(1 + AU + 2\sqrt{AU}) * \text{VSWR}}$$

P_{max} = derated maximum power
M = Amplitude modulation index (100% = 1)
AU = Aural to visual ratio (20% aural: AU = 0.2)

Much of the technical literature written on this topic discusses methods to increase the peak power rating through the use of increased pressure and high dielectric strength gases. While in theory a higher rating is possible, the system requires significant monitoring to prevent extensive damage due to breakdown if a failure in the pressurization integrity occurs. The decision to utilize these procedures requires extensive analysis of the proposed system and should not be used if other more stable configurations are available.

viewing various manufacturers' specifications. This is typically a result of a different safety factor used when deciding on the actual maximum frequency of operation. A 5–10% reduction in the calculated f_c is a normal safety factor and will account for manufacturing tolerances and the effects of connections and elbows. However, where numerous elbows assembled "back-to-back" will be used within a system, additional safety factors may be necessary to prevent the generation of higher order modes.

Voltage Standing Wave Ratio

Typically, it is important that the transfer of energy from the source (transmitter) to a load (antenna or receiver) be as efficient as possible. The attenuation constant is one factor in determining the efficiency as it represents a direct loss of energy. Another factor is the impedance mismatch between the line and the load. The greater the mismatch, the higher the reflected level of energy at the connection. This reflected energy reduces the amount of signal transmitted to the load and results in lower efficiency. The incident and reflected waves will combine to produce an uneven voltage distribution. Voltage and current maxima and minima occur resulting in a standing wave along the line. The VSWR is defined as the ratio of maximum to minimum voltage:

$$\mathrm{VSWR} = V_{max}/V_{min}$$

The effects of VSWR may be significant depending on the VSWR level and the specific conditions of operation. In general, as the VSWR increases, the maximum voltage levels increase and voltage breakdown may occur. Excessive heating along the line may also occur due to the current maxima and softening or melting of the dielectric insulating material may cause the inner conductor to shift in position.

In a simplified form, the impedance at any point along a transmission line is described as:

$$Z(s) = Z_o \times (1 + V^-/V^+)/(1 - V^-/V^+)$$

$Z(s)$ = resultant impedance at a point on the line
V^- = reflected voltage component
V^+ = forward voltage component
Z_o = characteristic impedance of the line
Z_L = impedance of the load

The reflection coefficient Γ_o at the load is defined as:

$$\Gamma_o = (Z_L - Z_o)/(Z_L + Z_o)$$

Then

$$\mathrm{VSWR} = V_{max}/V_{min} = (1 + \Gamma_o)/(1 - \Gamma_o)$$

or

$$\Gamma_o = (\mathrm{VSWR} - 1)/(\mathrm{VSWR} + 1)$$

Attenuation

Attenuation is defined as the loss created by the imperfect conductivity of the conductors and the imperfect insulating dielectric medium. In coaxial lines, losses come from both the inner and outer conductor materials and the material used to support the inner conductor. In solid dielectric cables, the dielectric loss can be appreciable and at higher frequencies actually exceed the conductor losses. For air dielectric lines, the insulating supports are a small percentage of the total dielectric space and their losses are generally negligible.

Attenuation and efficiency are directly related. Efficiency is defined as the ratio of power delivered to the load relative to the power input into the transmission line:

$$\text{Efficiency} = \text{Power out/Power in}$$

The efficiency is determined by calculating the total attenuation of the line based on its overall length. The total attenuation (in db) is then converted to efficiency:

$$\text{Efficiency \%} = 10^{-(\alpha_{total}/10)} * 100$$

Actual attenuation can be influenced by the VSWR of the line and the operating temperature. Except in cases of extreme VSWR or temperatures, these effects are insignificant to the performance of the system.

For VSWR, the loss is increased by the factor

$$\frac{1 + \mathrm{VSWR}^2}{2 * \mathrm{VSWR}}$$

For temperature, the greatest effect is a change in the temperature of the inner conductor since the conductivity of the material is affected by temperature. The adjustment factor for attenuation (M_a) is given by:

$$M_a = \sqrt{1 + \sigma_o(T_t - T_o)}$$

T_t = inner connector temperature °C
T_o = inner conductor temperature at standard rating °C
σ_o = temperature coefficient of resistance at standard rating

For standard temperature rating of 20°C, σ_o = 0.00393/°C. Then:

$$M_a = \sqrt{1 + 0.00393\,(T_t - 20)}$$

For an inner conductor temperature of 100°C, the attenuation will increase by a factor of 1.146.

Power Ratings

There are two primary ratings when discussing the power handling capabilities of coaxial transmission lines—average and peak power. The *average* power rating is based on the maximum internal heating the line construction can withstand. This is normally limited by the maximum safe operating temperature of the dielectric spacing material. The *peak* power rating is based on voltage breakdown characteristics.

Peak Power

Peak power is defined as the maximum RF power which can be reached in any interval, such as an RF in defined lengths, typically no longer than 20 feet. The individual lengths are then attached to each other through the use of flanges and the inner conductors are typically "spliced" together. Rigid coaxial lines can have diameters up to 14 inches. They have high power handling capabilities and low attenuation values.

Design Criteria: Electrical Parameter Optimization

A basic principle in the design of most transmission lines is to find the optimum configuration of inner conductor to outer conductor dimensions where the power handling is maximized and the attenuation minimized. Each configuration will then determine a characteristic impedance for the line.

Characteristic Impedance

The characteristic impedance is determined by the relative dimensions of the inner and outer conductors and the dielectric material between them. It can be expressed by the following equation:

$$Z_o = \frac{60}{\sqrt{\epsilon'}} * \ln\left(\frac{D}{d}\right)$$

where:

Z_o = characteristic impedance
ϵ' = dielectric constant or relative permittivity of dielectric to air
D = inside electrical diameter of outer conductor
d = outside electrical diameter of inner conductor

Attenuation

The attenuation of the line is normally expressed in terms of loss per unit length or db/100 ft (db/100 meters). The attenuation is due to dielectric losses and conductor losses. The dielectric material loss is directly proportional to frequency. For "air dielectric" lines where the majority of the volume of space between the inner and outer conductors is air, the most commonly used dielectric materials are TFE and polyethylene. These materials produce very small losses, and in the case of rigid coaxial lines, this loss is usually negligible relative to the conductor losses. Conductor losses are related to dimensions, permeability and conductivity of the material. The conductor loss varies with the square root of the frequency and for copper conductors:

$$\alpha = \frac{0.433}{Z_o} * \left(\frac{1}{D} + \frac{1}{d}\right) * \sqrt{f}$$

where

α = attenuation constant, db/100 ft
f = frequency, MHz

Attenuation is minimized when D/d is equal to 3.59, which results in an impedance of 77 Ω.

Power Handling

The power handling capabilities of coaxial lines are based on two factors: the maximum peak power (or maximum voltage that can be safely present) and the maximum average power, which is determined by the allowable temperature rise on the inner conductor.

Peak Power

The maximum electric field strength between two coaxial conductors can be calculated from:

$$E_{max} = \frac{0.278}{d} * \sqrt{\frac{P}{\ln\left(\frac{D}{d}\right)}}$$

where:

E_{max} = Maximum electrical field strength, volts/in
P = Power level of signal, watts

E_{max} is at a minimum when the ratio D/d is equal to 1.65. This results in a characteristic impedance for an air dielectric line of 30 Ω.

Determining the average power capability of a line requires extensive testing or the use of complex thermal models. It has been determined that the optimum ratio of D/d is approximately 2.72, resulting in a characteristic impedance of 60 Ω.

Summary

It can be seen from the previous analysis that there are tradeoffs between optimum configurations for attenuation, peak power and average power. Broadcast transmission lines have standardized on 50 and 75 Ω. The 50 Ω impedance is a balance between optimum peak power and average power. Typically, if power is the controlling factor, a 50 Ω line is used. If attenuation (efficiency) is more important, then a 75 Ω line should be used.

Design Criteria: Operational Parameters

With the characteristic impedances determined, the choice of line now is based on the desired frequency of operation (or limited by the cut-off frequency), and the actual power handling and attenuation ratings.

Cut-off Frequency

The cut-off frequency (f_c), is the frequency above which undesirable modes of propagation can be generated. The generation of these modes results in degraded efficiency, higher reflections (Voltage standing wave ratio—VSWR) and lowered power handling. Since larger coaxial transmission lines can handle more power, the (f_c) typically is the limiting factor where frequency of operation versus power handling is determined. The cut-off frequency is inversely proportional to the conductor dimensions and is determined by the following equation:

$$f_c\,(GHz) = \frac{7.52}{\sqrt{\epsilon'} * (D + d)}$$

Differences in the maximum operating frequency of specific line sizes are sometimes evident when re-

2.4
COAXIAL TRANSMISSION LINES

KERRY W. COZAD
ANDREW CORPORATION, ORLAND PARK, IL

INTRODUCTION

Transmission lines are one of the main components in the RF transmission plant of a broadcast station. Acting as the connecting link between the transmitter and the antenna, the transmission line plays a critical role in both the quality and reliability of the broadcast signal. Therefore, the proper choice of a transmission line type to be used can have a significant impact on the success of the station.

The choice of transmission line is typically decided based on the following criteria:

- Frequency of Operation
- Power Handling
- Attenuation (or efficiency)
- Characteristic Impedance
- Tower Loading (size and weight)

With the implementation of additional digital broadcast channels, other criteria such as installation costs, connector design, reliability and the effectiveness of transmitting multiple channels are receiving renewed attention from design and broadcast engineers. There are a wide variety of transmission line types and designs to choose from. This chapter reviews the attributes of the various types and describes performance comparisons to assist the broadcast engineer in the selection of the type best suited for a specific need.

COAXIAL TRANSMISSION LINE TYPES

A coaxial transmission line consists of two concentric conductors, the inner conductor being supported within the outer conductor through the use of a dielectric material (see Figure 2.4-1). The dielectric material may be continuous throughout the line or, as in the case of rigid coaxial lines, located at distinct points along the line in the shapes of "pegs" or cylindrical "beads." Some general characteristics from electromagnetic field theory are:

- An infinite number of electromagnetic field configurations (modes) are possible
- Propagating modes will be in a general form of waves traveling along the axis of the line
- The propagation constant is different for each mode
- There is a frequency called the cutoff frequency where the propagation constant is 0. Below this frequency, there is no propagation of that mode.

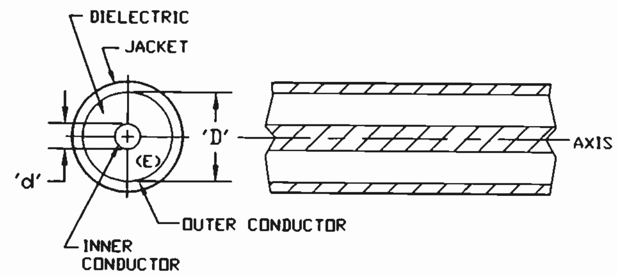


Figure 2.4-1. Coaxial cable cross section.

- There is one mode, transverse electromagnetic mode (TEM) for which the electromagnetic fields are transverse to the transmission line axis. The TEM has a cut-off frequency of 0 Hz. This is the mode that is primarily used for propagating signals within coaxial transmission lines.

Based on these common characteristics, coaxial transmission lines are usually divided into three primary groups: flexible (RF) cables, semiflexible cables and rigid coaxial lines.

RF cables are typically used for short interconnections between equipment. They consist of a solid or stranded inner conductor, solid plastic dielectric insulating material and a braided outer conductor sheath. It is coated with a plastic outer jacket to resist moisture and abrasion. Typically having outer conductor diameters of less than 0.5 inch, its attenuation characteristics restrict it to short runs and because it is highly susceptible to deterioration due to aging, it is normally only used indoors.

Semiflexible cables are designed with soft tempered copper inner and outer conductors. The dielectric material may be either solid (foam filled) or a spiral that has been wrapped in helical fashion around the inner conductor (air dielectric). Again, a plastic jacket is applied to the outer conductor to resist abrasion. One advantage of semiflexible cables is that they can be fabricated in diameters up to 9 inches and in continuous lengths of hundreds and even thousands of feet. This cable type is used extensively for radio and low power television broadcasting, as well as interelement feeders for some antenna types.

Rigid coaxial lines are designed with hard tempered inner and outer conductors. Discrete dielectric insulators are used to support the inner conductor within the outer conductor. Because it is rigid, it must be fabricated

Surge Protectors Used in Low-Voltage AC Power Circuits.
15. *ANSI/IEEE C62.41 Guide on Surge Testing for Equipment Connected to Low Voltage AC Power Circuits.*
16. *ANSI/IEEE C62.45 Guide on Surge Testing for Equipment Connected to Low Voltage AC Power Circuits.*
17. *IEEE 142 (Green Book) Grounding Practices for Electrical Systems.*
18. *IEEE Standard 518 Recommended Guide on Electrical Noise.*
19. *UL-1283 Standard for Safety—Electromagnetic Interference Filters.*
20. *UL-1449 Standard for Safety—Transient Voltage Surge Suppressors.*
21. *NFPA 70 National Electrical Code.*
22. *NFPA 75 Standard for Protection of Electronic Computer Systems.*
23. *NFPA 780 Standard for the Installation of Lightning Protection Systems.*
24. *Military Standard* (MIL Std.) 220A.
25. Federal Information Processing Standards (FIPS) Publication 94; CCITT Rec. K-17 *Waveform Specification for Electronic Systems.*
26. *NEMA LS-1 Low Voltage Surge Protective Devices.*
27. *IEC 61643-1, International Electrotechnical Committee Standard for Surge. Protection Devices Connected to Low Voltage Power Distribution Systems. Part 1: Performance Requirements and Testing Methods.* (1998.)
28. Williams, Earle R. *The Electrification of Thunderstorms*, Scientific American, November 1988.

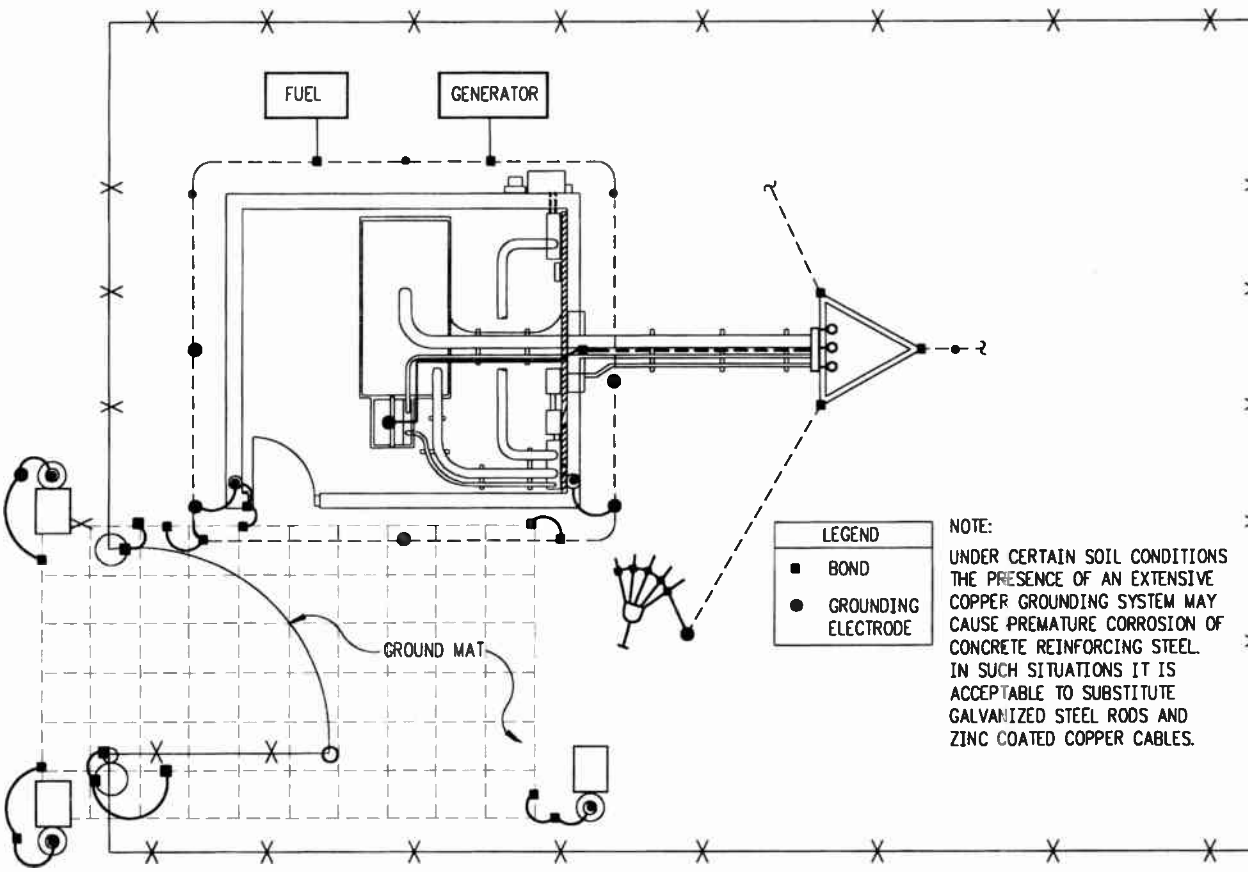


Figure 2.3-26. Typical composite bonding & grounding installation.

CONCLUSION

This chapter has been written in tutorial form as every site is different and no single set of recommendations will apply to every situation. The principles set forth, while tailored to a broadcast environment, apply equally to other systems. There are still a few mysteries to be solved in completely understanding lightning, but once it enters a wiring system, it becomes an electrical current which is both predicable and understandable.

REFERENCES

1. United States National Lightning Detection Network? 1989–1996 Average U.S. measured lightning flash density (flashes/km^2/year).
2. Military Handbook 419A, *Grounding, Bonding and Shielding for Electronic Systems*. December, 1987. (Available from the U.S. Government Printing Office, Philadelphia, PA).
3. Cianos & Pierce, *A Ground Lightning Environment*, Courtesy SRI International (Standard Research Institute), Menlo Park, CA.
4. *Lightning Protection Using Transzorbs*, General Semiconductor Corporation.
5. FCC Rules and Regulations, Chapter 68, section 3.02—*Lightning Protection for Telecommunications Equipment.*
6. *NFPA 780—Lightning Protection Code*, National Protection Association, Quincy, MA.
7. *UL-96A—Standard for Lightning Protection Systems*, Underwriters Laboratories, Northbrook, IL.
8. *LPI-175—Standard of Practice*, Lightning Protection Institute, Woodstock, IL.
9. Portions adapted from Military Handbook 419A listed in reference 2.
10. Stromberg Carlson Practices, Section 53-001-03, *DCO-CS Grounding Engineering Considerations, 1989.*
11. Telecommunication Electrical Protection—AT&T, 1985.
12. *ANSI/IEEE C62.33 Standard Test Specifications for Varistor Surge Protective Devices.*
13. *ANSI/IEEE C62.35 Standard Test Specification for Avalanche Junction Semiconductor Surge Protective Devices.*
14. *ANSI/IEEE C62.36 Standard Test Methods for*

- All surge suppressors should be UL 1449 tested, approved and labeled
- Ask for published proof of testing to back up claims, especially clamping levels, energy handling capabilities (joules, watts or current) and speed of response
- Look for 3 to 5 years unconditional warranties
- Review the 6 major performance characteristics: Response time, voltage protection level, power dissipation, disturbance free operation, reliability and operating life. Make sure maximum power dissipation level and voltage protection level using specific waveforms, are stated at the same point to avoid misinterpretation
- Talk to other similar users for advice and recommendations
- Consult industry standards for performance requirements. At present, there is no testing or application standard for coaxial surge suppressors, so a comparative review of reputable manufacturer data will be necessary. Compare: impedance (50 or 75 Ω); clamping speed; frequency range; number of transmit, transceive or receive only signals; transmit power; presence of ac or dc power with the RF signals; mounting; connector type and sex.

Isolated Ground Receptacle

Figure 2.3-25 shows an isolated ground receptacle circuit commonly used in computer room grounding applications. The receptacles used in this type of circuit differ from the norm in that their ground pins are electrically isolated from their mounting tabs. They are therefore isolated from their outlet box and structural ground at each receptacle location. A dual system of grounding conductors ensures that equipment plugged into an isolated ground (IG) receptacle references ground first at the single point ground.

The use of isolated ground receptacles helps to ensure that plug-in terminals, printers, diagnostic and other ancillary equipment is properly referenced to the single point ground and not the local structure. It only takes one item of equipment connected between the protected equipment and a remotely grounded receptacle to compromise the integrity of the grounding system.

Figure 2.3-26 is a composite of the bonding and grounding recommendations for the typical broadcast site. While complicated in appearance, each component has its purpose as part of a simple-to-understand subsystem.

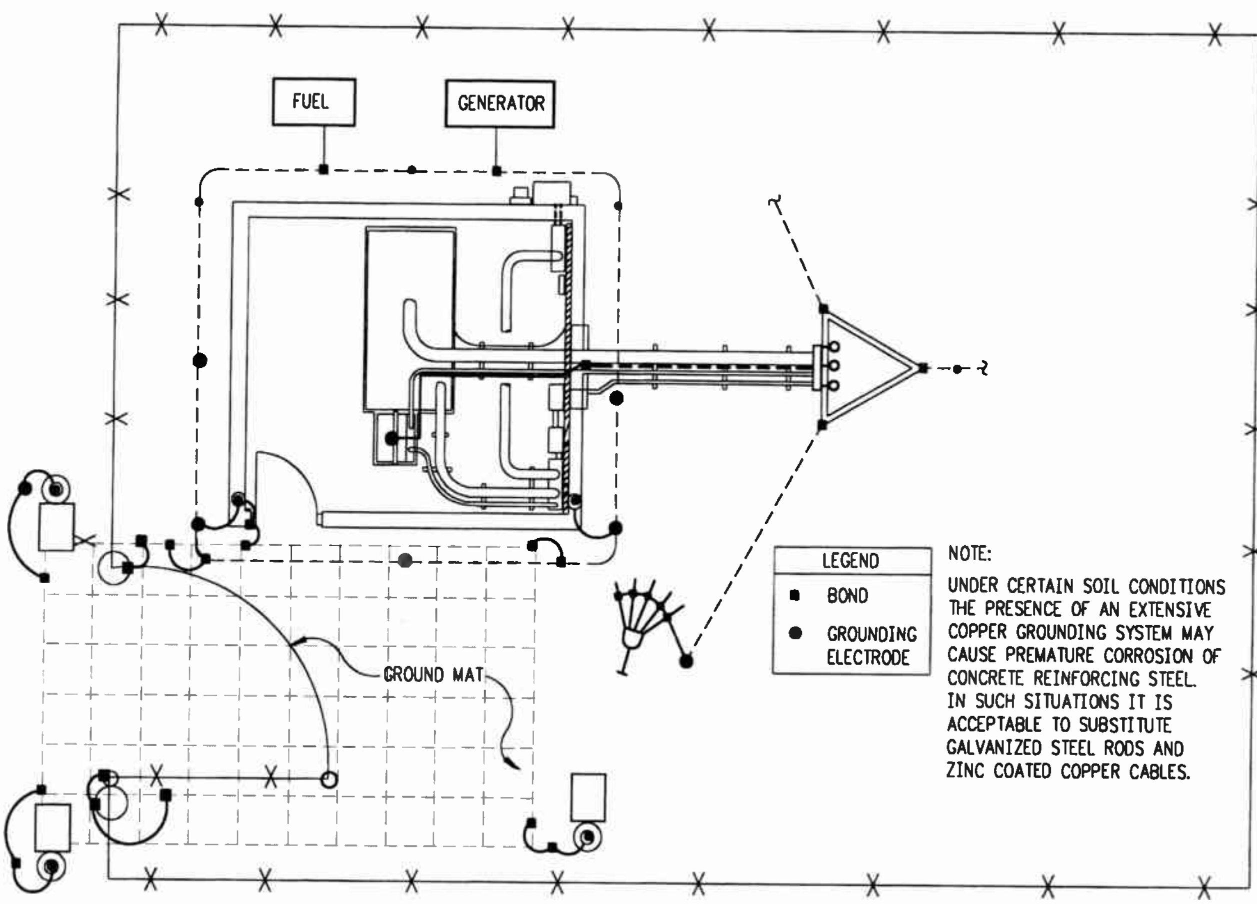


Figure 2.3-25. Isolated ground receptacle.

open during servicing and each condition can cause equipment flashover.

Flashover within electrical equipment is serious in itself, however, the problem is compounded when operating voltage is also present. The flashover arc provides a low impedance path for 60 Hz follow current and significant damage to the equipment may occur. The arrester will prevent this condition from occurring by limiting the travelling wave voltage.

Signal Line Suppressors

Figure 2.3-24 shows a common surge suppressor configuration for telephone signal line and coax applications. Most suppressors of this type are both shunt type and multi-stage hybrid devices utilizing a high-energy first stage, a fast acting second stage and impedance in series between the two stages to coordinate their clamping behavior. Because of the multi-stage design, these devices must be installed in series with the protected circuits.

The treatment of shields is often an issue when dealing with signal line surge suppressors. Figure 24 shows shields being bonded to the suppressor ground bus to force them to track the single point ground. If ground loop or other technical restrictions prevent direct bonding of the shields, they should be protected with suppressors as any other active circuit.

The bonding lead distance for signal line suppressors is often more critical than for power devices. Tolerance levels of signal circuits are normally lower than for power supply inputs, and the relatively small voltages developed in the suppressor ground leads can be become significant.

Signal line and coaxial surge suppressors are sold in a variety of shapes and sizes for different applications. The suppressors should generally clamp transient voltage on a circuit to within 150% of normal peak operating voltages and even lower in some applications. Since most signal line suppressors are inserted in series with the circuit being protected, it is wise to evaluate the effect of their series impedance and capacitance on the insertion loss of a circuit.

The effect of suppressor capacitance can be important in many high speed data, RF and video applications. One simple way of evaluating the effect of this capacitance is to equate it to equivalent cable feet. For example, if the desired suppressor exhibits capacitance of 100 picofarads, and the cable used in the circuit is rated at ten picofarads per foot. Will the circuit tolerate an additional 10 ft of cable? If so, the suppressor capacitance should produce no noticeable effect on the circuit.

Surge Suppressor Selection

Surge suppressors and lightning arresters come in all shapes, sizes, types and accessories. Contacting surge suppressor manufacturers may be extremely confusing due to the myriad of choices available, but eventually a choice must be made. Contacting the chairmen of the referenced IEEE, U.L. and IEC standard committees may also be helpful in understanding application issues. The following guidelines therefore may be helpful:

- Select suppressors and arresters based on IEEE C62.41 Categories (A, B or C)

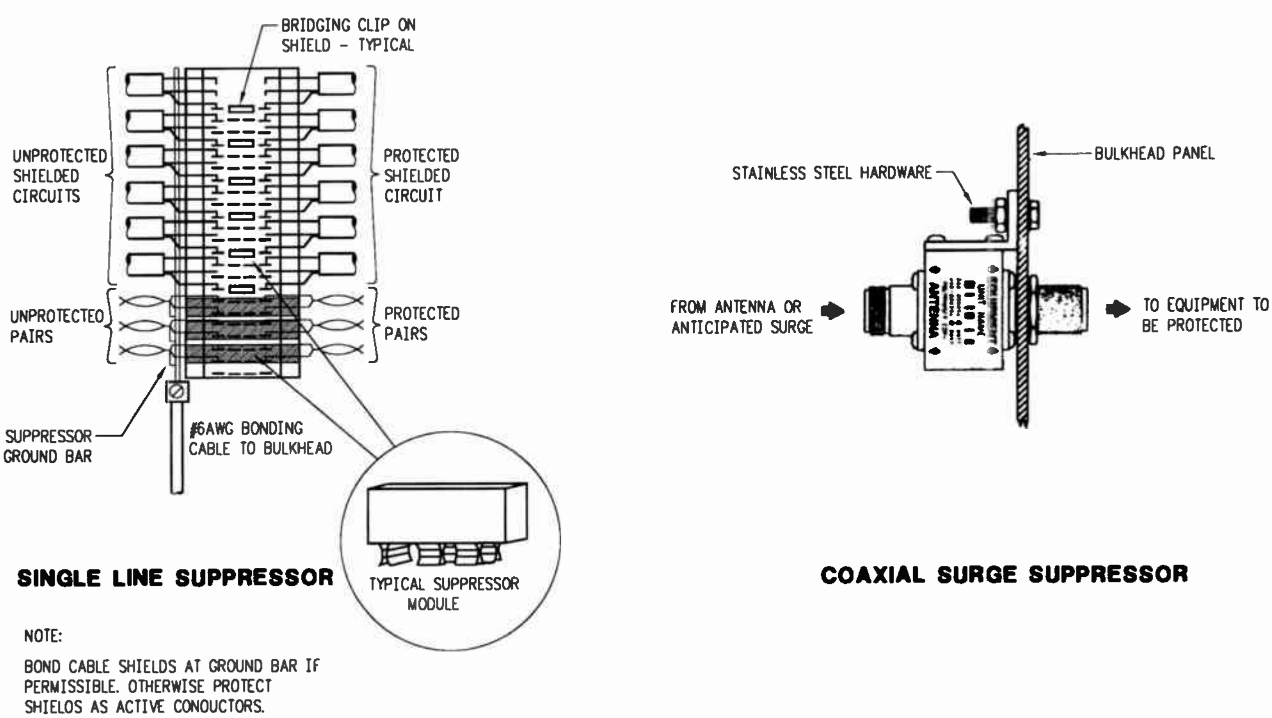


Figure 2.3-24. Common signal & coaxial suppressors.

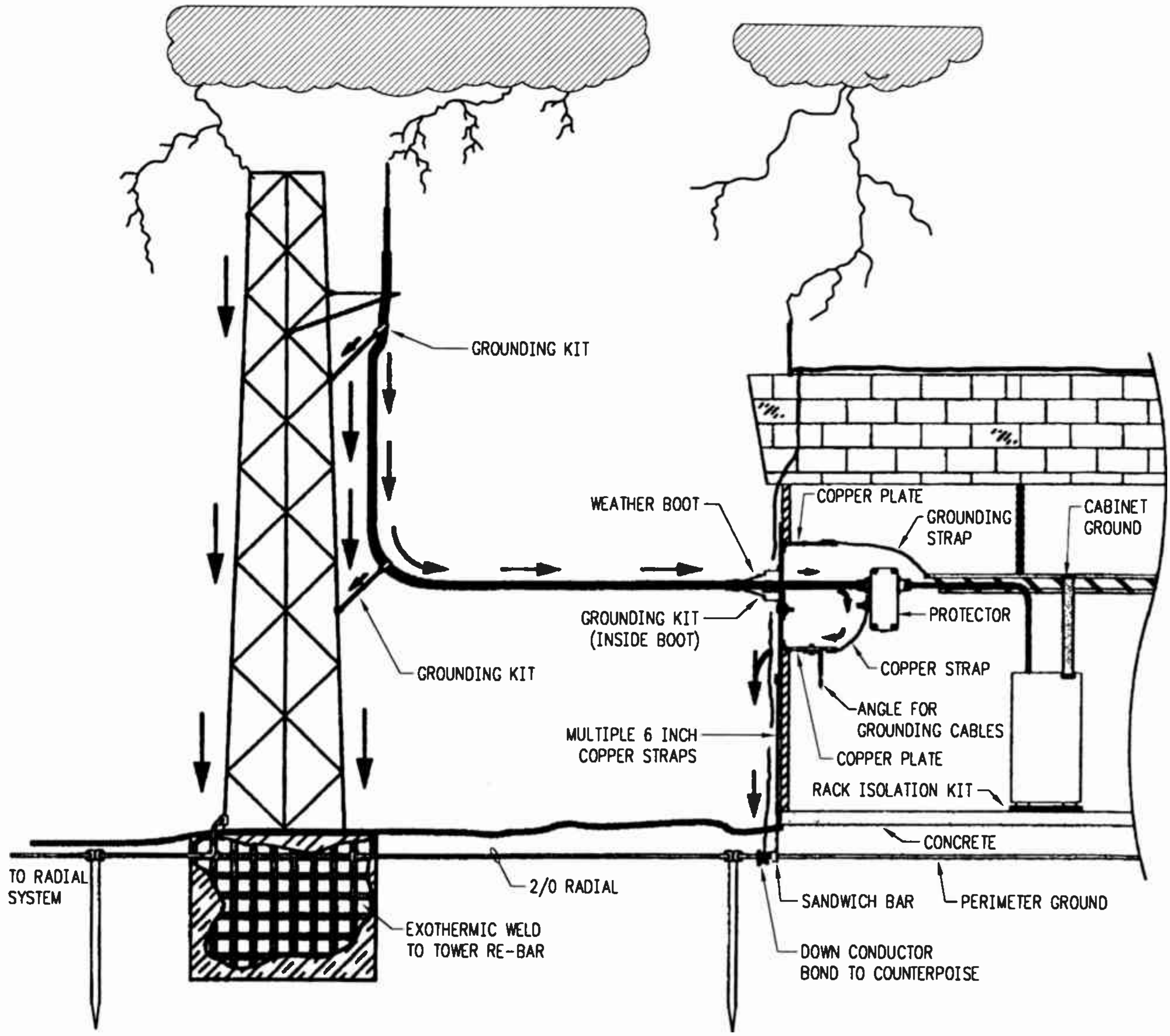


Figure 2.3-23. Typical bulkhead installation with single point grounding.

Surge Suppression

In Figure 2.3-22, surge suppressors are shown at the transfer switch emergency feed to protect the transfer switch from surges on the emergency power feeder whenever the feeder and generator starting controls are not run in rigid metal conduit. Where metal conduit is used, it must be bonded to the perimeter ground ring. Main service surge suppressors should be connected through individual or integral fused switches or breakers with short circuit interrupting ratings equal to or higher than the available service fault current. Also, a suppressor is installed on the circuit for tower lights. Both suppressors should be bonded to the grounding bus, keeping the length of their bonding lead as short as possible. Suppression should also be provided on all metallic, control, monitoring and communications lines and circuits serving lighting, winches and other electrical items outside the building. Again, locate these suppressors at the point of entry for the circuits and bond their ground leads to the bus with the shortest possible lead length.

The lightning arrester shown at the main disconnect outside the building serves an important purpose. Under normal operation, the main disconnect is closed and the transfer switch is connected to utility power. During operation on emergency power the normal transfer switch position is open. Potentially, a lightning strike to the utility line will propagate along the line as a travelling wave in both directions from the point of lightning contact. Upon reaching the open circuit input of the transfer switch, the wave will reflect back on itself, potentially doubling its initial crest value. The same condition can occur at the main utility company disconnect if it should be plex may be engineered with a single point grounding system and proper surge suppression on its external circuits. The terminals and their printers, however, are scattered throughout the building, referencing ground at each location through their power cords. It is possible to designate the ground pin on the receptacle for each terminal/printer combination as the single point ground for the equipment at that location. A combination power and data suppressor may be provided for each location which ensures that these conductors are held within safe limits of the receptacle ground pin and chassis of the equipment. The equipment is isolated from stray grounds by placement on a desk top.

Figures 2.3-22 and 2.3-23 are examples of how single point grounding applies to a typical broadcast tower-building installation. A bulkhead panel in the wall of the building serves as a single point ground reference for all equipment within the small facility. All coaxial cables, waveguides, and raceways from the tower are bonded to the bulkhead as they pass into the building.

The physical size of the electrical equipment dictates that it cannot be located directly at the bulkhead panel. To minimize the effect of bonding system inductance, a 6 in. wide bonding bus is extended to each side of the bulkhead or for large installations an overhead ground ring is provided (width of the strip provides the necessary low inductance). Its 1/4 in. thickness, while not necessary for electrical reasons, provides the installer with a bus which may be drilled and tapped to accept short bonding pigtails to the equipment.

Support hangers for cables and raceways serving the protected equipment are isolated from the roof structure to prevent inadvertent current flow through the raceways. Isolation is provided between the equipment feet and floor slab by a high dielectric polypropylene pad. Such isolation may not be necessary if adequate isolation is provided by the equipment feet. Nylon bolts may be used with conventional expansion anchors to secure equipment to the floor without violating the integrity of the single point grounding system.

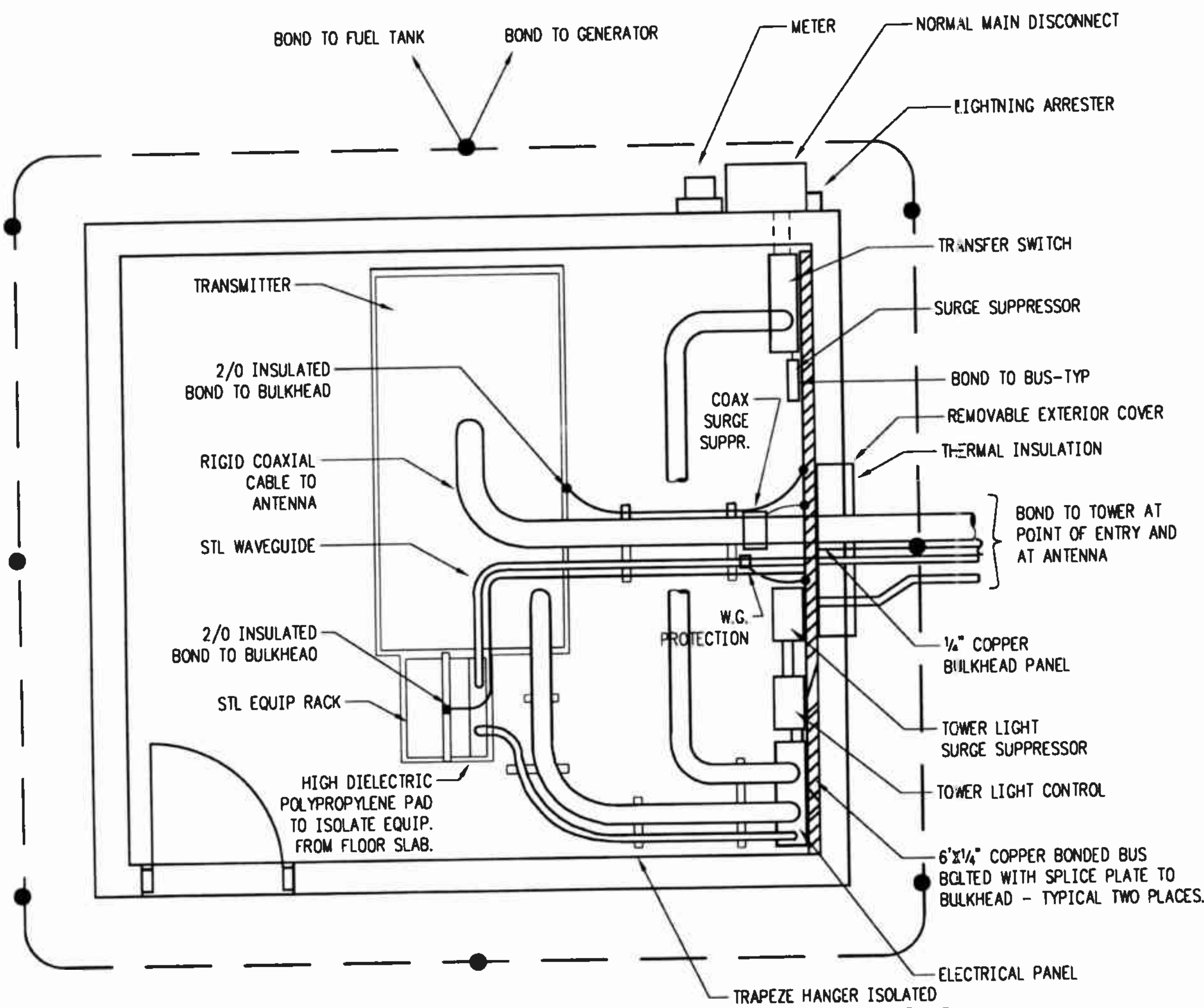


Figure 2.3-22. Single point transmitter building grounding.

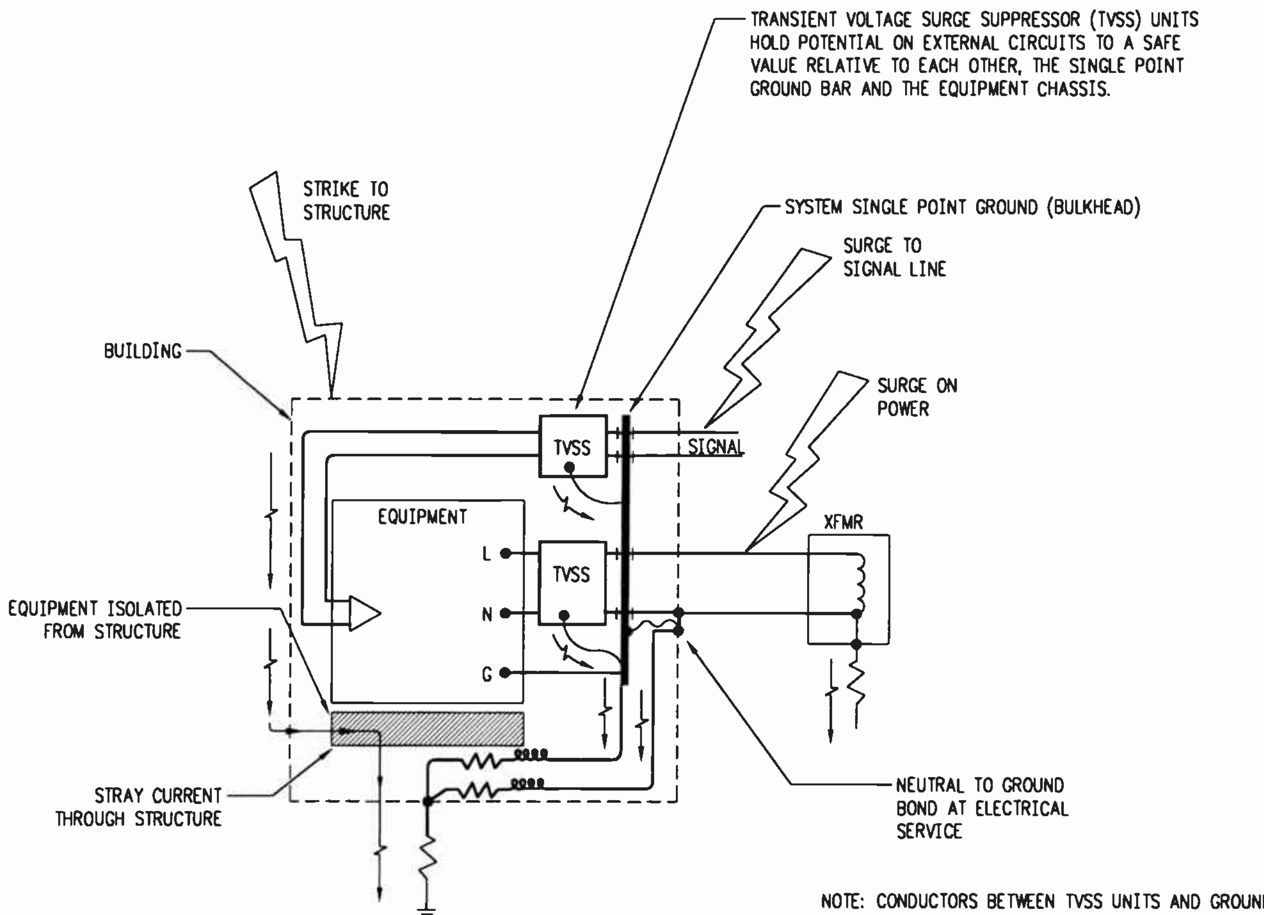


Figure 2.3-21. Practical application of single point grounding with TVSS.

chassis and all external metallic circuits are tied together at the single point ground. The only difficulty with the example in Figure 2.3-20 is that *nothing works*!

Figure 2.3-21 provides a more realistic approach to single point grounding. The only difference between Figure 2.3-21 and the preceding example is that Transient Voltage Surge Suppression (TVSS) devices are used on the active circuits, which, for obvious reasons, cannot be directly bonded to the single point ground.

For the purpose of this discussion, it is helpful to think of surge suppressors as a conditional bond, clamping or limiting the excursion of voltage on active circuits to a safe level relative to each other and to the single point ground. The single point ground may rise and fall in potential as the suppressors discharge current into it or during a strike to the building, but the difference in potential presented to the protected equipment is always held within safe limits.

It is worthwhile noting that TVSS devices will clamp in response to a rise in potential on their ground terminal as well as for legitimate transients on their active conductors. A strike to the building or nearby structure will cause a significant elevation in ground potential. The single point ground will rise in potential by virtue of its connection(s) to the building grounding system. The TVSS units, seeing their building ground terminals rise in potential above their remotely connected active circuits, will clamp, forcing the active circuits to track the potential of the single point ground and the chassis of the protected equipment. Again, the voltage excursion seen by the equipment is held to a safe level and no damage is sustained.

Applications of Single Point Grounding

The application of single point grounding is normally limited to equipment within a room or a group of rooms. While it is possible to design larger configurations, the need to bring circuits in at different locations soon dictates the need for multiple locations, each treated as an island of equipment with its own suppression devices and single point of ground reference. Larger single point grounding systems are also more susceptible to induced voltages from nearby lightning by virtue of their increased cable lengths.

As an example of two extremes in scale, a computer room which serves terminals throughout a station com-thickness is mechanical strength and mounting convenience.

As a final note on the inductance of bonding materials, *never expect an insulated conductor in steel conduit to carry lightning current effectively.* Ferrite beads make a reasonable effective low pass filter when placed around an insulated conductor. Steel conduit around an insulated cable creates the same effect, increasing the inductance of the cable within the conduit at least an order of magnitude similar to a choke coil. Where this condition exists (and there are many locations) a marginal compromise is to bond both ends of the cable to the conduit, permitting the conduit to serve as part of the circuit.

Single Point Grounding

If, after the exercise in bonding, you conclude there is no way of preventing potential differences in a conductor carrying lightning current, you are correct. There is, however, a method of preventing lightning current flow through a bonding circuit. With no current flow there can be no potential difference between the bonded items. This method is called *single point grounding.*

In Figure 2.3-20, the equipment chassis and all metallic circuits leaving the equipment for the outside world have been bonded together and to a ground conductor at a single point. There is no possibility of a difference in potential between the circuits entering the equipment or between these circuits and the chassis as they are all bonded together. There is also no possibility of current flow from the single point ground into the equipment through any of the circuits as the equipment is isolated from the structure.

A surge entering on the power or signal lines cannot present itself to the equipment in differential mode as the lines are all connected together. A common mode surge arriving at the single point ground will pass harmlessly to ground through the grounding conductor and ground electrode resistance. There will be potential rise at the single point ground due to inductance and resistance in the grounding circuit, however; no current can flow through the equipment as it remains isolated from other points of ground reference. The equipment will simply rise and fall in potential, tracking the potential of the single point ground.

A lightning strike to the building or other structures connected to its grounding system will also cause the single point ground to rise and fall in potential. The equipment, however, sees no potential difference as its

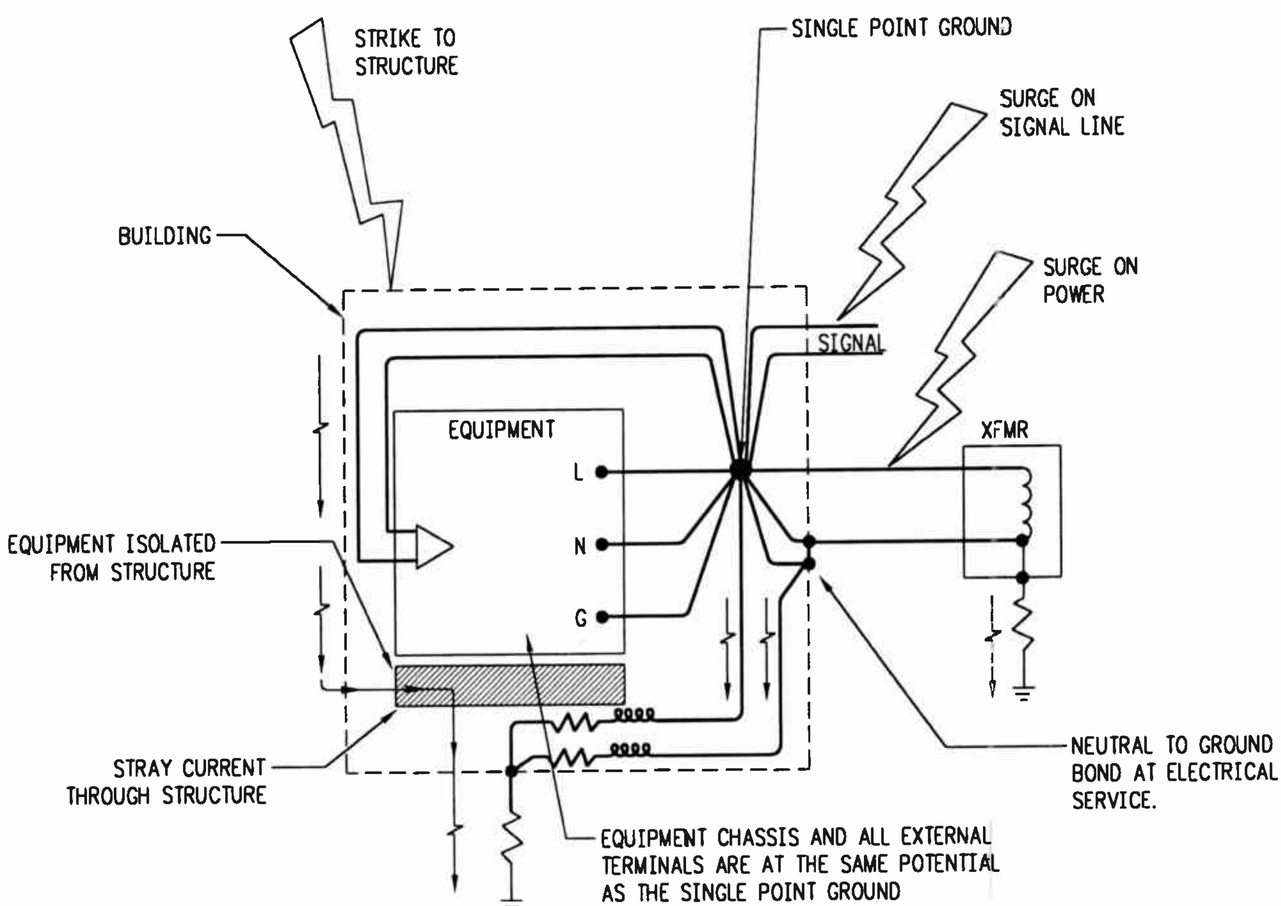


Figure 2.3-20. Single point grounding concept.

Kirchoff's law states that current will divide itself among all of the available parallel paths through a circuit in proportion to the impedance of each path. Lightning currents behave in the same way, flowing through all available paths to ground. In direct current circuits, the voltage produced across any circuit component is the product of current and resistance. When dealing with rapidly changing lightning current, inductance of the circuit plays a far larger role than simple resistance. Recall that an inductor tends to oppose any change in current until it has stabilized its magnetic field.

The bonding conductor shown in Figure 2.3-18 by virtue of its connection to ground at remote building A, serves as one of several paths for lightning current to follow on its way to ground. In this case, only 5% or 1,000 amperes of the lightning current flows through the 75 ft conductor with the remainder of the 20,000 A flowing into the grounding system of building B and its electrical service. Assuming a 1.5 μs risetime for the lightning current, the peak end-to-end voltage on the conductor is 15,030 V 15,000 V of the total is the result of inductance in the conductor. The remaining 30 V is the result of the conductor's resistance.

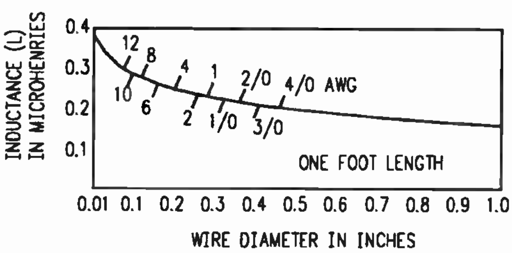
The normal reaction to lowering the potential difference between buildings in this example is to suggest a larger cable. After all, larger cables have less resistance and lower voltage drop. Changing to a larger cable, however, has little effect on the circuit inductance, affecting primarily the 30 V portion of the total.

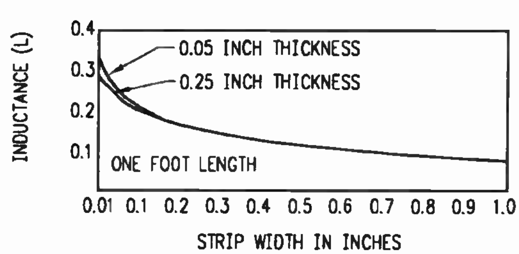
Figure 2.3-19 provides a comparison of inductance values for a 1 ft. length of various sizes of round conductors, strip materials and coaxial cables. Strip materials are considerably more effective for the same cross sectional areas as a round conductor, making them more attractive as a bonding medium for lightning protection purposes and bulk head grounding. Another interesting property of strip material is that it reaches a width of about 4 in., the major reason for

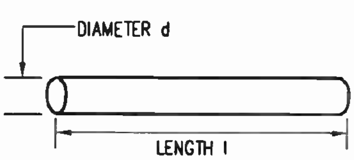
COAX DIAMETER

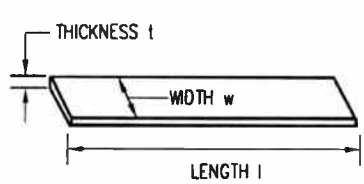
LENGTH IN FEET	½"	⅞"	1"	1⅝"	2"	3"
100	51.0	48.0	47.0	44.2	43.0	40.4
150	81.0	76.0	74.4	70.0	68.0	64.3
200	111.0	104.0	102.0	97.0	94.2	89.2
300	174.0	164.0	161.4	152.5	148.7	141.2
500	305.0	289.0	284.5	270.0	263.4	251.0

COAXIAL CABLE









L(in uH) = 0.5081 [2.303 $\log_{10}$ (4 L/d) − 0.75
x 10^{-2}

WHERE l=LENGTH IN in.
d=DIAMETER IN in.

ROUND CONDUCTOR

L(in uH) = 0.5081 [2.303 $\log_{10}$ (2 L/w +t) + 0.5
+ 0.2205 (w + t/L] x 10^{-2}

WHERE l=LENGTH IN in.
w=WIDTHR IN in.
t=THICKNESS IN in.

RECTANGULAR SHAPE

Figure 2.3-19. Inductance values for common engineering materials.

Differential or normal mode surges are often more damaging than their common mode counterpart as most equipment is designed to operate in a differential fashion. In power circuits, a differential mode surge may appear on one or more phases relative to the neutral and ground conductors. In signal circuits, especially those operating on a balanced differential basis, the tolerance to differential mode surges is lower than for common mode. Common mode surges on electrical power systems are routinely converted to differential mode at electrical services where one side of the service is referenced to ground. The same conversion process can occur on a balanced circuit when an upstream transient voltage surge suppressor (TVSS) device clamps one side of a pair before the other.

Ground differential damage is a bit more obscure than either of the other mechanisms, however, it is responsible for a great deal of damage to systems with equipment in multiple locations. Equipment in different buildings, or even equipment within different areas of the same building can be damaged through ground differentials.

In the simple example of Figure 2.3-17, assume that building B receives a lightning strike of 20,000 amperes. Also assume the grounding system resistance is 2 Ω. As the lightning current flows into the earth through the 2 Ω grounding resistance, a 40,000 V potential rise will be produced in the building B grounding system. Since the equipment in building B references the local building through its power cord and bonding conductor, its chassis will rise to about the same potential.

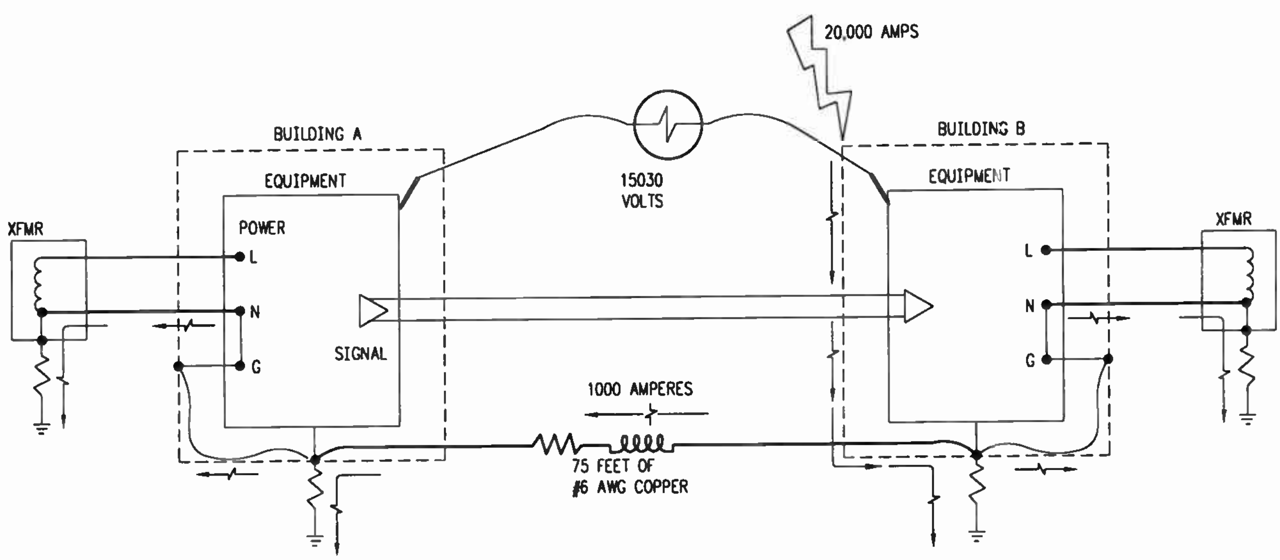
Circuitry within the building B equipment will attempt to track the building ground potential rise, except for the components which attach to wiring from building A. These components see a large difference in potential between the balance of their circuitry and the wiring to building A. Building A has not been involved in the ground potential rise so these circuits are still near ground potential.

Component breakdown occurs within the equipment in building B and a small fraction of the total lightning current attempts to find a path to ground through the wiring leaving for building A. Upon reaching the equipment at the remote building A, this current presents itself as a common mode surge causing damage to the equipment.

Equipment Protection Strategies

Protecting equipment from the effects of lightning involves a combination of grounding, bonding and surge suppression. Grounding provides a path to introduce lightning currents into the earth. Bonding serves to equalize lightning potential differences between various elements of equipment. Surge suppression limits differences in potential on active circuits which cannot be directly bonded.

Bonding is a means of equalizing potentials during a lightning strike. Figure 2.3-18 repeats the earlier example in Figure 2.3-17 regarding damage due to ground potential differences. This time, however, a bonding conductor is provided between the two grounding systems in an attempt to keep both at the same potential.



I = 1000 AMPERES
R = .03 9/16¢m3/4 (FROM WIRE TABLES)
L = 22.5 MICROHENRIES (0.3uh/ft 75 FEET)
di = 1000 AMPRES (SAME AS I)
dt = 1,5 MICROSECONDS RISETIME

$$E = IR = L\frac{di}{dt}$$

$$E = 1X10^{3} X\ 310^{-2} = 22.5X10^{-6} X\ \frac{1X10^{2}}{15X10^{-6}}$$

$$E = 30 = 15000$$

$$E = 15030 \text{ VOLTS}$$

Figure 2.3-18. Example of ground potential difference.

nearly equal potentials at both ends, additional bonding conductors between the radials will carry little or no current. There is only a slight advantage in bonding between radials for guy wire currents introduced into the grounding system at guy anchor points.

Figure 2.3-11 shows a perimeter ground ring or counterpoise which encircles the station building helping to equalize potential differences within the building. The station building ring also serves as a connection point for driven electrodes, fencing and other objects which must be bonded. A bonding conductor is shown between the station building ground ring and the tower base or grounding system. This conductor will equalize potentials and minimize the level of current carried between the tower and station building by the coaxial lines.

A commercially available bulkhead plate is shown on the side wall of the station building and is bonded to the station grounding ring. This plate serves as a single point ground for all equipment within the station building and the coax lines. In new construction, the steel reinforcing mesh in the station building floor should be bonded together to the bulkhead panel to minimize potential differences between the equipment and floor during a lightning strike.

EQUIPMENT DAMAGE PROCESS

Most lightning damage to equipment occurs as the result of potential differences which exceed the tolerance level of the equipment. These potential differences may be presented to the equipment or system through external metallic circuits as a conducted current or induced transient voltage surge. They may also occur as the result of differences in ground potential at various items of equipment which are connected together to form a system. (Figure 2.3-17 describes these situations in greater detail.)

Common mode surges, which are also referred to as *longitudinal mode*, arrive at the equipment with approximately equal potential on both sides of a balanced pair or on a number of circuits simultaneously. These surges may be induced into a wiring system by nearby lightning, directly coupled into the circuit or even created by the action of an upstream suppression device as it clamps two or more conductors together. Common mode surges may enter equipment on power phases, signal and other circuits. Damage is normally sustained due to potential differences between the affected circuit(s) and equipment chassis or other uninvolved circuits.

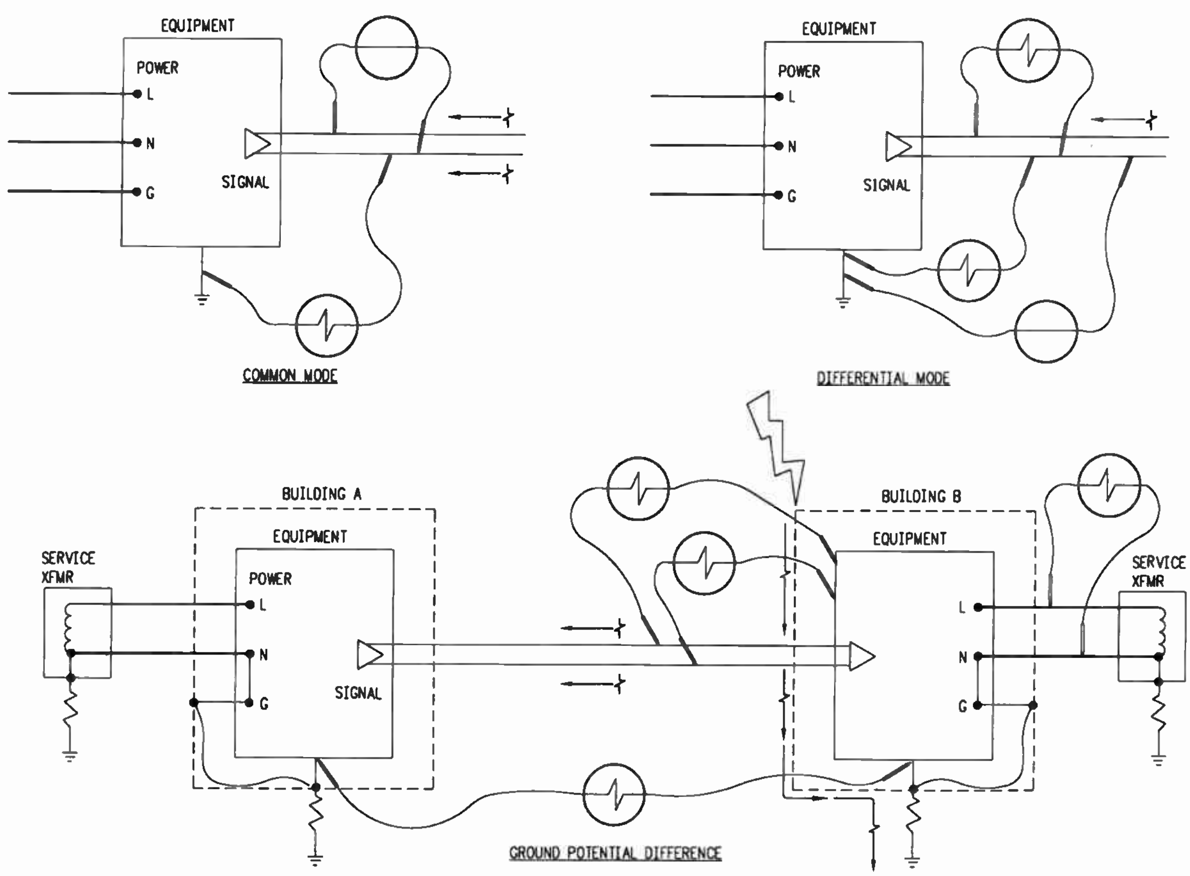


Figure 2.3-17. Typical equipment damage process.

current flowing between the tower and station building grounding systems, reducing potential differences between the two systems. The potential rise on the tower due to a lightning strike is a result of fast rising lightning current, tower inductance and ground resistance such that $E = IR + L\,di/dt$ where E is the potential rise from top of tower to ground, I is the magnitude of lightning current, R is the grounding resistance, L is the tower inductance and di/dt is the lightning current rise with respect to time. Tower structures have a certain amount of inductance per foot. The amount of this inductance is dependent upon the geometric configuration as well as the width of the tower. This width-to-height ratio will determine the total inductance of a tower. A 150 ft tower for instance, with 35 in. side widths can have an inductance of 40 μH. This value of inductance can be approximated ($W/H \leq 1\%$) by treating the tower as a 1/4 wave antenna using:

$$f = \frac{468 \times 10^6}{2(\text{H in feet})} \text{ then inductance } L = \frac{377}{2\pi f}$$

Therefore, there is not much that can be done about the tower inductance once the tower structure is known. Grounding, however, can be improved with proper grounding techniques.

Figure 2.3-16 describes a typical grounding configuration for a guyed tower and associated transmitter building. Tower grounding may be accomplished either by a system of interconnected driven electrodes between the tower base and guy anchor points or by radial counterpoise conductors without rods. The chosen method should be determined from previously described electrode calculations and earth resistivity testing. In the driven electrode configuration, one radial counterpoise conductor is extended from the base of the tower to each guy anchor point. This conductor interconnects driven electrodes near the tower base, at guy anchor locations and at intermediate points from any wire currents.

The rings shown feeding the system of radial conductors should stop within a few feet of the tower base. The complex mesh created by multiple bonds between the rings and radials will help to feed lightning current efficiently from the tower legs into each of the radial conductors. Apart from providing more copper in contact with the earth, there is no advantage in adding additional rings in the area between the tower base and guy anchor points. Current flowing from the tower base out on the radials will produce approximately equal potentials between adjacent radials. With

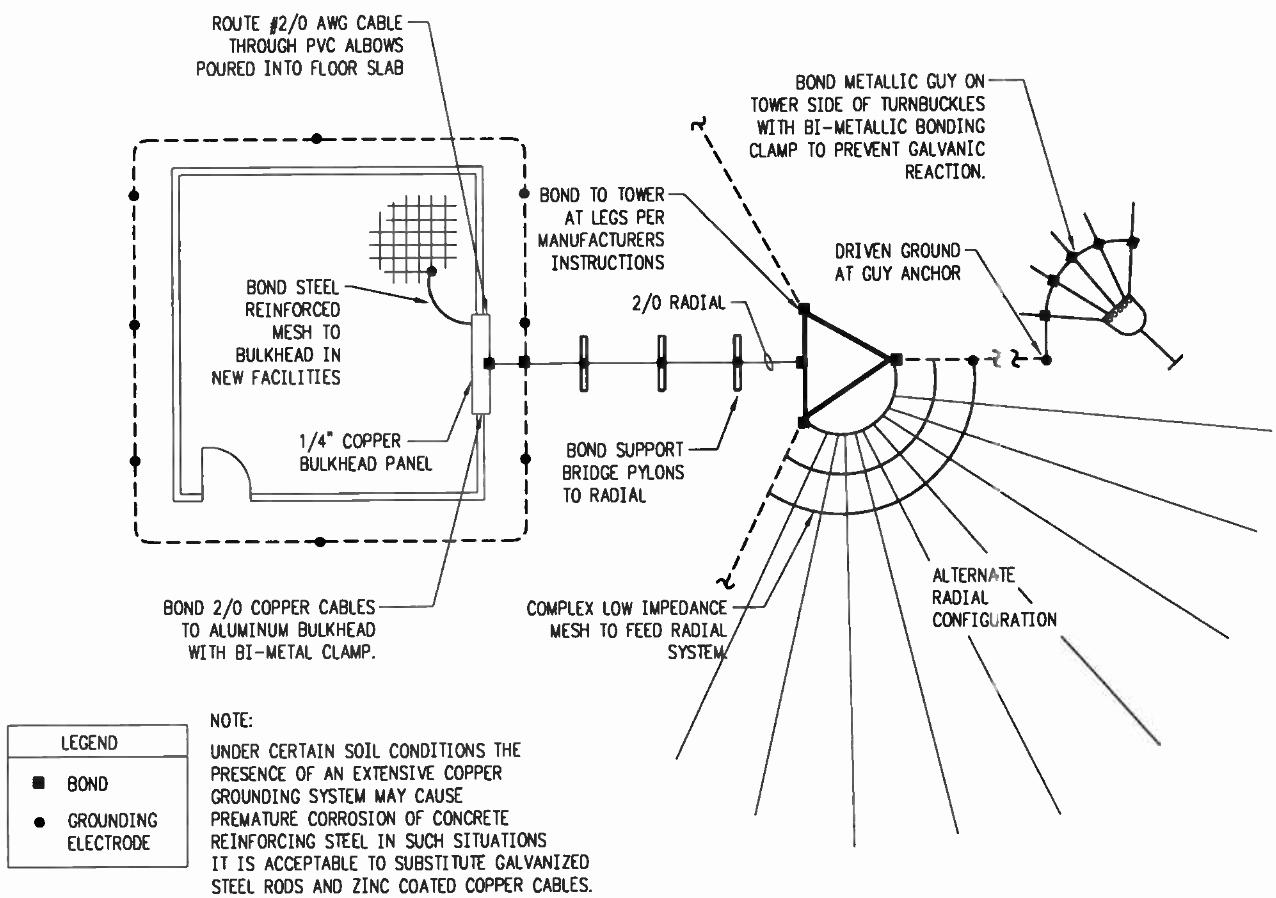


Figure 2.3-16. Typical tower ground system.

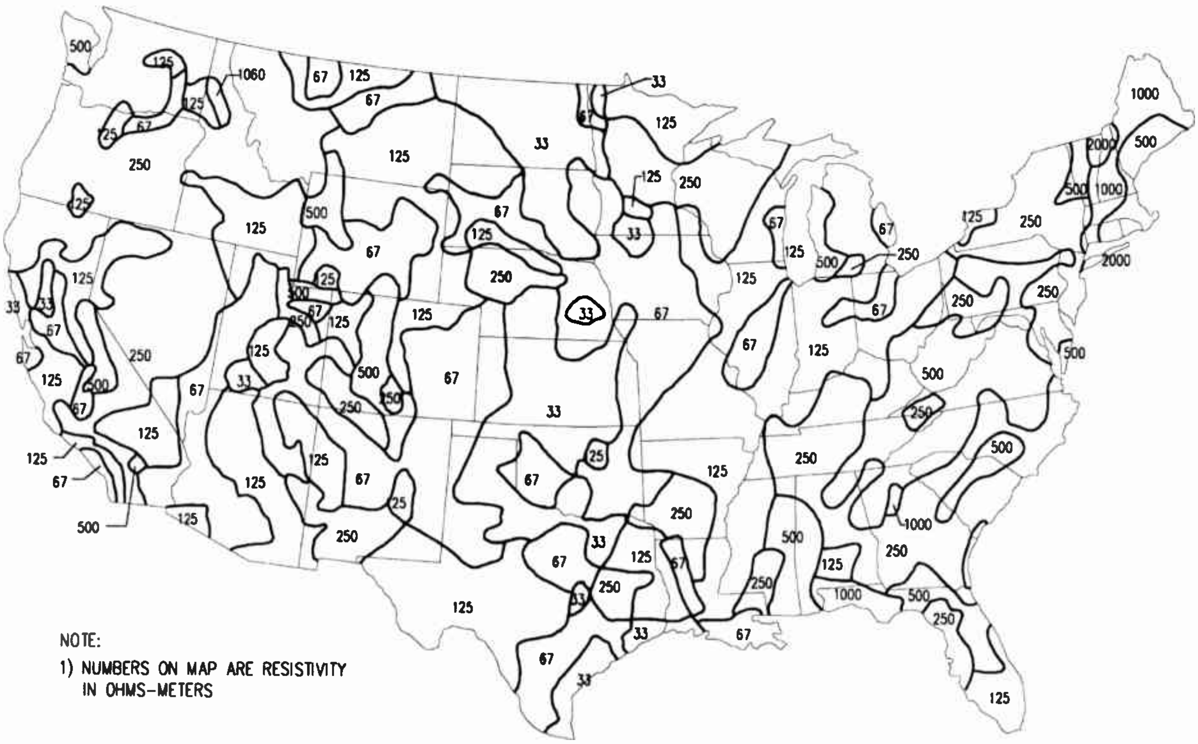


Figure 2.3-14. Estimated average earth resistivity in U.S.

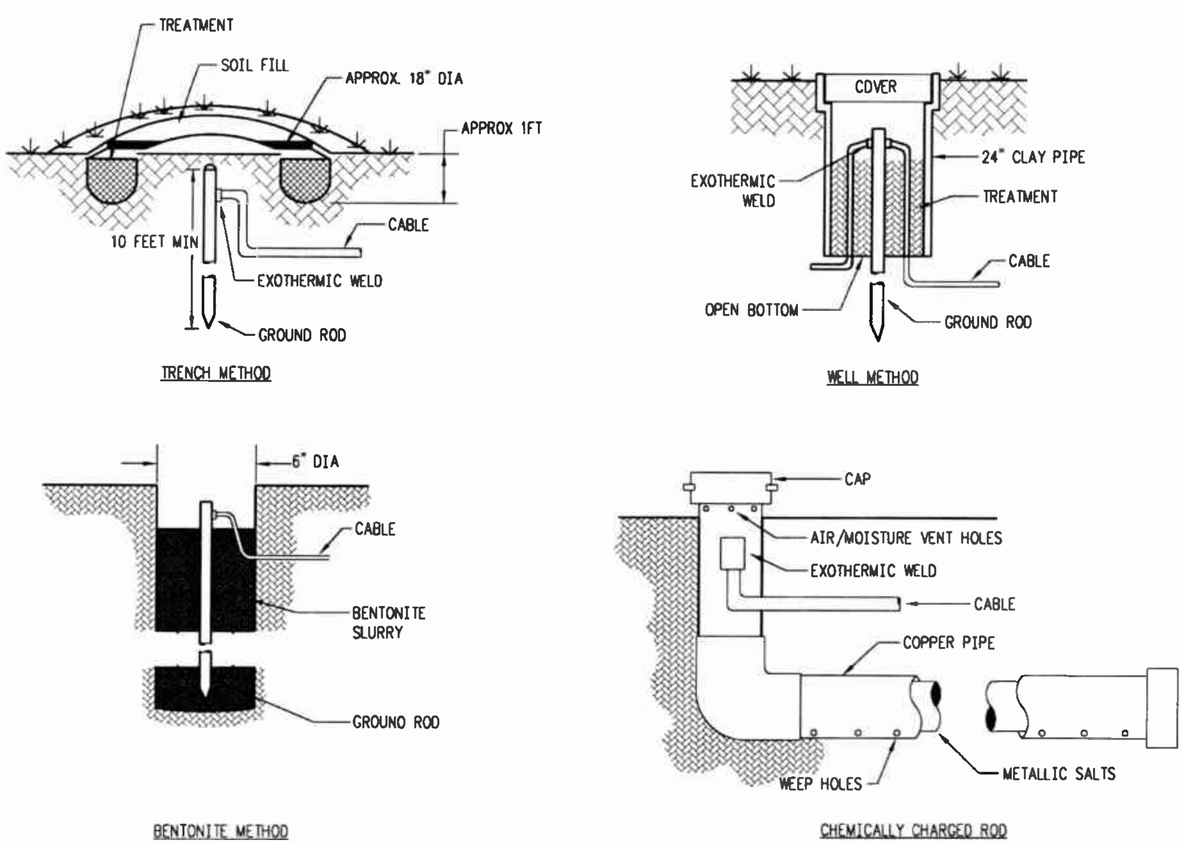


Figure 2.3-15. Common chemically treated ground systems.

Formulas for Calculation of Resistances to Ground

	Hemisphere radius a	$R = \frac{\rho}{2\pi a}$
	One ground rod length L, radius a	$R = \frac{\rho}{2\pi L}\left(\ln\frac{4L}{a} - 1\right)$
	Two ground rods $s > L$; spacing s	$R = \frac{\rho}{4\pi L}\left(\ln\frac{4L}{a} - 1\right) + \frac{\rho}{4\pi s}\left(1 - \frac{L^2}{3s^2} + \frac{2L^4}{5s^4}\right)$
	Two ground rods $s < L$; spacing s	$R = \frac{\rho}{4\pi L}\left(\ln\frac{4L}{a} + \ln\frac{4L}{s} - 2 + \frac{s}{2L} - \frac{s^2}{16L^2} + \frac{s^4}{512L^4}\cdots\right)$
	Buried horizontal wire length $2L$, depth $s/2$	$R = \frac{\rho}{4\pi L}\left(\ln\frac{4L}{a} + \ln\frac{4L}{s} - 2 + \frac{s}{2L} - \frac{s^2}{16L^2} + \frac{s^4}{512L^4}\cdots\right)$
	Right-angle turn of wire length of arm L, depth $s/2$	$R = \frac{\rho}{4\pi L}\left(\ln\frac{2L}{s} + \ln\frac{2L}{s} - 0.2373 + 0.2146\frac{s}{L} + 0.1035\frac{s^2}{L^2} - 0.0424\frac{s^4}{L^4}\cdots\right)$
	Three-point star length of arm L, depth $s/2$	$R = \frac{\rho}{16\pi L}\left(\ln\frac{2L}{s} + \ln\frac{2L}{s} + 1.071 - 0.209\frac{s}{L} + 0.238\frac{s^2}{L^2} - 0.054\frac{2}{L^4}\cdots\right)$
	Four-point star length of arm L, depth $s/2$	$R = \frac{\rho}{8\pi L}\left(\ln\frac{2L}{a} + \ln\frac{2L}{s} + 2.912 - 1.071\frac{s}{L} + 0.645\frac{s^2}{L^2} - 0.145\frac{s^4}{L^4}\cdots\right)$
	Six-point star length of arm L, depth $s/2$	$R = \frac{\rho}{12\pi L}\left(\ln\frac{2L}{a} + \ln\frac{2L}{s} + 6.851 - 3.128\frac{s}{L} + 1.758\frac{s^2}{L^2} - 0.490\frac{s^4}{L^4}\cdots\right)$
	Eight-point star length of arm L, depth $s/2$	$R = \frac{\rho}{16\pi L}\left(\ln\frac{2L}{a} + \ln\frac{2L}{s} + 10.98 - 5.51\frac{s}{L} + 3.26\frac{s^2}{L^2} - 1.17\frac{s^4}{L^4}\cdots\right)$
	Ring of Wire-diameter of ring D, diameter of wire d, depth $s/2$	$R = \frac{\rho}{2\pi^2 D}\left(\ln\frac{8D}{d} + \ln\frac{4D}{s}\right)$
	Buried horizontal strip length $2/L$, section $\propto$ by b, depth $s/2$, $b < a/8$	$R = \frac{\rho}{4\pi L}\left(\ln\frac{4L}{a} + \frac{a^2 - \pi ab}{2(a-b)^2} + \ln\frac{4L}{S} - 1 + \frac{s}{2L} - \frac{s^2}{16L^2} + \frac{s^4}{512L^4}\cdots\right)$
	Buried horizontal round plate, radius a, depth $s/2$	$R = \frac{\rho}{8a} + \frac{\rho}{4\pi s}\left(1 - \frac{7}{12}\frac{a^2}{s^2} + \frac{33}{40}\frac{a^4}{s^4}\cdots\right)$
	Buried vertical round plate, radius a, depth $s/2$	$R = \frac{\rho}{8a} + \frac{\rho}{4\pi s}\left(1 + \frac{7}{24}\frac{a^2}{s^2} + \frac{99}{320}\frac{99}{320}\frac{a^4}{s^4}\cdots\right)$

Notes:

1. Approximate formulas, including the effect of images
2. Dimensions must be in centimeters to return result in ohms
3. ρ = resistivity of earth in ohm-centimeters
4. For 10 ft. (3 m) rods of $\frac{1}{2}''$ (12.7 mm), $\frac{5}{8}''$ (15.88 mm) and $\frac{3}{4}''$ (19.05 mm) diameters, the grounding resistance may be quickly determined by dividing the soil resistivity ρ in ohm-centimeters by 292, 302 and 311 respectively.
5. Data source IEE Green Book (Std. 142-1982)

Figure 2.3-13. Formulas for calculation of resistances to ground.

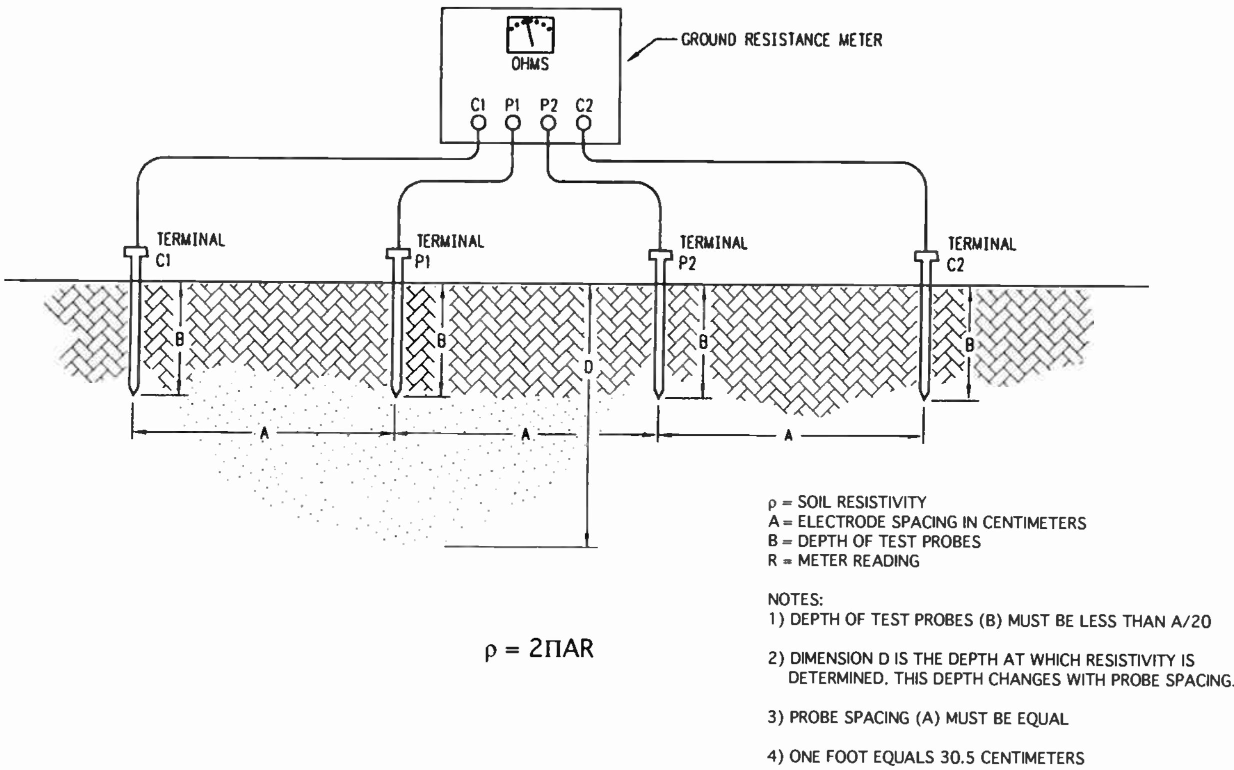


Figure 2.3-12. Soil resistivity test method.

frost line. They are often the only practical solution in areas with shallow subsurface rock. There have been instances at bald rock mountain top sites where radials were either grouted into saw cuts in the rock or simply pinned against the face of the rock.

The performance of a grounding system in high resistivity soil can often be improved through the addition of chemical salts. These salts leach into the soil, increasing the number of free ions with a proportionate decrease of soil resistivity in the area of the rod. Magnesium sulphate (Epsom salts), copper sulphate (blue virol), calcium chloride, sodium chloride (table salt), and potassium nitrate have been used for this purpose.

Figure 2.3-15 describes the trench and well methods for applying chemical treatment. A typical precharged chemical ground rod installation is also shown. In a precharged rod, moisture from the air enters the rod through breather holes at the top of the rod and leaches through chemicals inside, gradually exiting the rod through weep holes. As one might expect, chemically enriched grounds require recharging after a number of years to maintain their effectiveness. It is also wise to check with governing environmental agencies before introducing any foreign chemical into the soil.

The bentonite ground shown in Figure 2.3-15 is a popular way of decreasing grounding system resistance of a rod electrode in uniform soil by up to 30%. Instead of driving the rod, it is placed in the center of a 6 to 12 in. augured hole. A slurry consisting of bentonite clay and water (well drillers mud) is then poured around the rod. As the water settles out, the resulting clay remains moist through absorption of moisture from the surrounding soil. Popular additives to the slurry include up to 75% powdered gypsum (calcium sulphate) and up to 5% sodium sulphate (galvanic anode backfill).

Ground Electrode Testing

Testing of all grounding electrodes before they are connected to form a complex network is a fairly simple process that is well described in the documentation included with all ground electrode meters. The system as a whole, should be tested after all interconnections are made, providing a benchmark for future tests.

On a new site, it is often possible to perform ground system tests before the power company ground/neutral conductor is attached to the system. It is worthwhile to conduct a before and after test with probes in the same position to determine the influence of the power company attachment during future tests. It is also worthwhile to install permanent electrodes and marker monuments at the original P2 and C2 probe positions to ensure the repeatability of future tests.

Tower Building System

In typical tower building arrangements, the tower is normally subject to more frequent and larger lightning currents than the station building. It is therefore reasonable to place emphasis on the tower grounding system with less emphasis on that for the station building. Improved grounding at the tower will result in less

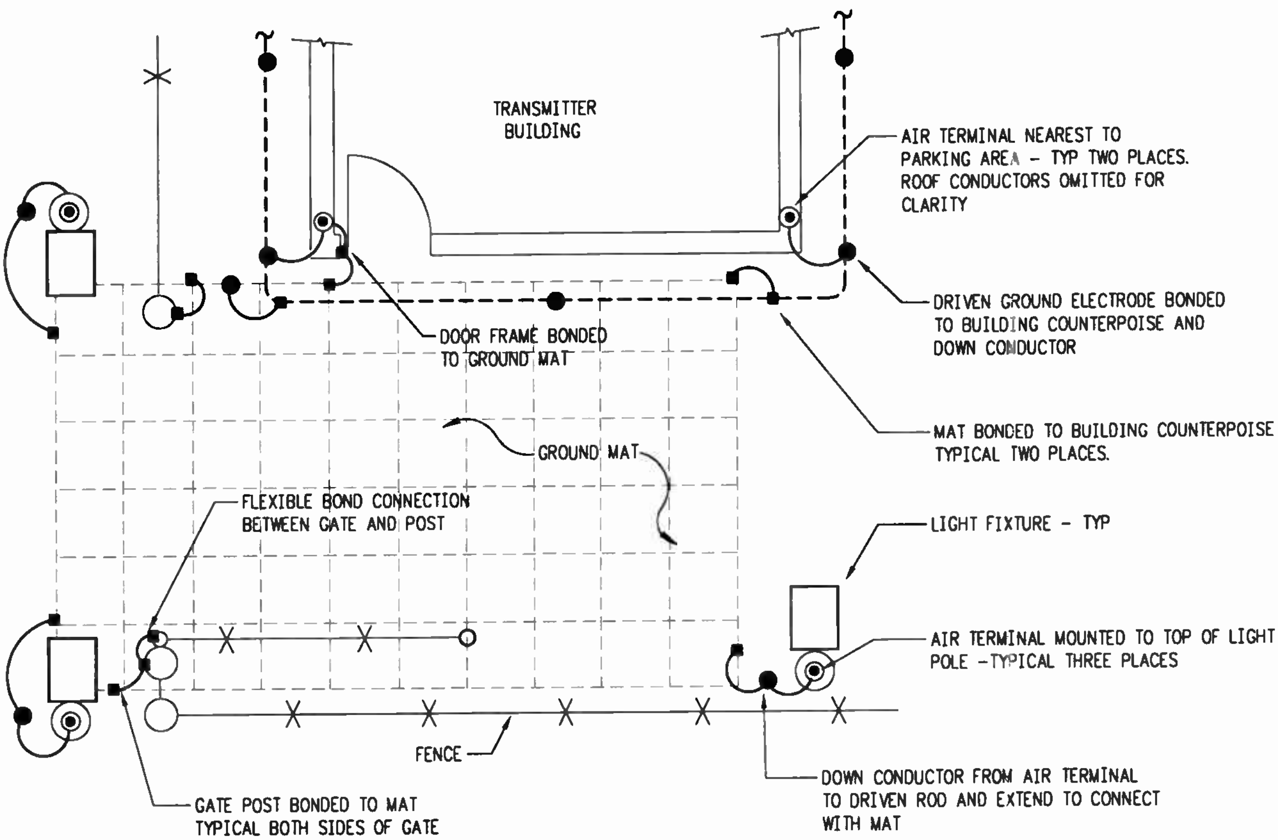


Figure 2.3-11. Personnel protection methods.

thermically welding a short length of flexible 2/0 AWG welding cable between the two elements.

The External Ground System

The effectiveness of a grounding system is a function of the type and extent of the electrode system used and resistivity of the surrounding soil. Soil resistivity is dependent on the quantity of free ions (chemical salts) in the soil, temperature and moisture content. The character of the soil below a particular site may also vary significantly with depth and location due to layering of different types of soil, the presence of hardpan layers, and subsurface rock.

Temperature is a major concern in shallow grounding systems because it has a major effect on soil resistivity. During winter months, the grounding system resistance may rise to unacceptable levels due to freezing of liquid water in the soil. The same shallow grounding system may also suffer from high resistance in the summer as moisture is evaporated from soil. It is wise to determine the natural frost line and moisture profile for an area before attempting design of a grounding system.

Figure 2.3-12 describes a four-point method for in place measurement of soil resistivity. Four uniformly spaced probes are placed in a linear arrangement and connected to a ground resistance test meter. An alternating current (at a frequency other than 60 Hz) is passed between the two most distant probes resulting in a potential difference between the center potential probes. The meter display in ohms of resistance may then be applied to the formula (shown in Figure 2.3-12) to determine the average soil resistivity in ohm-centimeters for the hemispherical area between the C1 and P2 probes.

Soil resistivity measurements should be repeated at a number of locations to establish a resistivity profile for the site. The depth of measurement may be controlled by varying the spacing between the probes. The probe length should not exceed 20% of the spacing between probes.

Once the soil resistivity for a site is known, calculations can be made to determine the effectiveness of a variety of grounding system configurations. Figure 2.3-13 presents equations for several driven rod and radial cable configurations which, after the solid resistivity is known, may be used for the purpose of estimating total system resistance. Generally, driven rod systems are appropriate where soil resistivity continues to improve with depth or where temperature extremes indicate seasonal frozen or dry soil conditions. (See Figure 2.3-14 for a typical U.S. soil resistivity map.)

Radials are also quite effective if placed below the

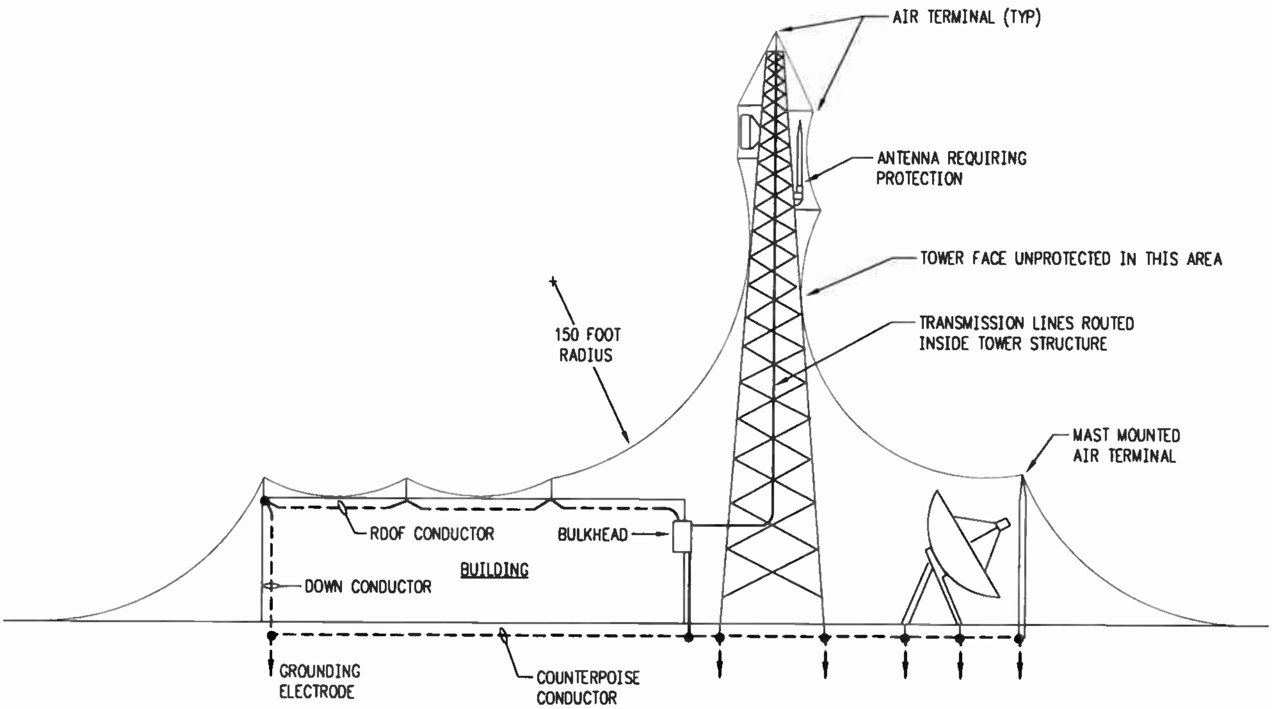


Figure 2.3-10. Example of *rolling ball theory* application.

Whenever practical, coaxial transmission lines and other wiring should be extended up the inside face of the tower to minimize the possibility of direct lightning contact.

Personnel Protection

The threat to personnel during a lightning strike ranges from the obvious danger of direct contact with a lightning strike to the more obscure effects of step and touch voltages. Protection from a direct strike when near or within structures is accomplished with traditional rolling sphere concept methods. However, to ensure that an adequate protective zone is provided in areas frequented by personnel the rolling sphere should be reduced to 100 ft radius.

Step and touch potentials are created as a lightning current passes through resistive soil and other available paths as it dissipates into the earth. A person in contact with only one point of the gradient will simply rise and fall in potential with the gradient without injury. A person in contact with multiple points on the earth or objects at different potentials along the gradient will become part of the current path and may sustain injury or death.

Figure 2.3-11 indicates a number of methods for protecting personnel from the direct and secondary effects of lightning.[2,9,17] A typical tower/transmitter site is used as an example. A technician responding to a service problem during a thunderstorm would likely exit his vehicle outside the gate, unlock and open the gate and move his vehicle into the inside yard. the technician would then leave the vehicle and enter the building.

The threat of a direct lightning strike to the technician has been minimized by establishing a protective zone over the areas to be traversed. This zone is created by the tower and air terminals mounted atop light poles.

Step potentials are minimized through the use of a ground mat buried just below the surface of the area where the technician is expected to be outside the vehicle. Ground mats are commercially available fabricated in 6 × 6 in. square pattern using #8 AWG bare copper wire. Each intersection is welded creating, for all practical purposes, an equipotential plan which short-circuits the step potential gradient in the area above the mat. The mat, as a whole, will rise and fall in potential due to lightning current discharges, however, there will be little difference in potential between the technician's feet. Mats should be covered with 6 in. of crushed stone or pavement.

The threat of dangerous touch potentials is minimized by bonding the ground mat to the building perimeter counterpoise, the fence at each side of the gate opening, bonding to the door frame of the transmitter building door, and providing a flexible bonding connection between the swing gate and its terminal post. Such bonding ensures that the object being touched by the technician is at or near the same potential as their feet.

Bonding both sides of the gate opening to the mat helps to ensure that the technician and both sides of the gate are at approximately the same potential while the gate is being handled. The flexible bond between the gate and its support post may be accomplished using a commercially available kit or by exo-

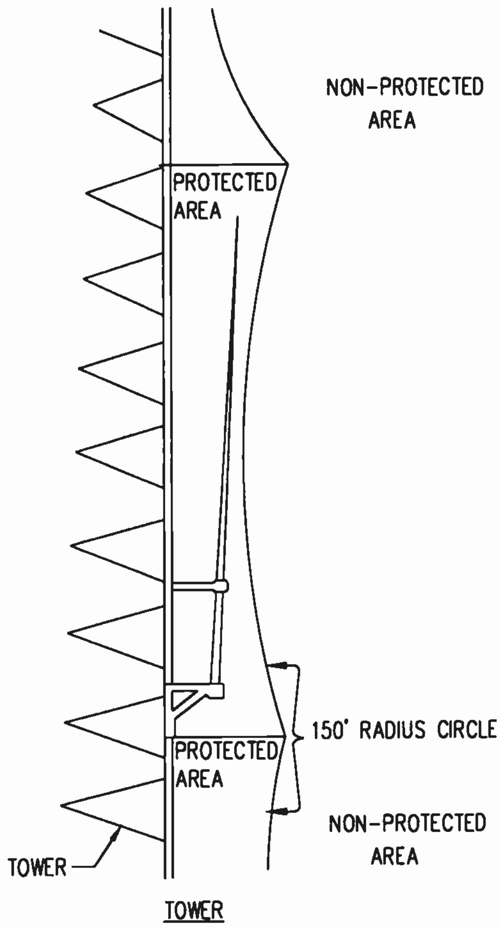


Figure 2.3-9. Side antenna protection method.

circuited in rigid metal galvanized conduit and the fixture itself should be constructed of metal to minimize surge currents on the wiring that would be difficult and expensive to eliminate from the source panelboard in the adjacent building. Surge suppression on the lighting circuit would still be required at the panelboard but it could be selected to handle minimal surge current.

A separate lightning protection *down conductor* is not necessary for adequate dissipation of lightning currents. The tower itself will always conduct the bulk of any lightning strike although a portion will travel on the coax shield and warning light conduit. An additional down conductor that is attached to the tower at intervals required by code will have a negligible effect on the division of current paths due to the inductance distribution of all components.

Lightning currents traveling down towers can, over time, cause arcing at tower joints, stripping away galvanizing and resulting in rusting. Maintenance, therefore of the tower system is important to long lasting, well performing components. Air terminals subject to heavy lightning currents of long duration can cause pitting and melting of points as well as severe mechanical stresses in mounting hardware. Bonding straps, cable fasteners and other miscellaneous hardware can loosen and allow movement or *whipping* of conductors from electromagnetic forces associated with unequal down currents in various components of the tower system. At least yearly maintenance is therefore recommended.

For guyed towers, guy wires can be very effective in dividing the lightning currents into additional paths to ground, therefore lowering the current and voltage stress to tower components and the service building bulkhead. Where guy wires contain insulators, special precautions must be observed. The insulators will be subject to voltage differences during lightning strikes occurring between guy to tower attachment points and the ground potential at the anchors due to ground currents and induced guy wire voltages. These voltage differences can reach hundreds of thousands of volts and can create arcing or *tracking* over insulators, resulting in ineffective isolation over time. Therefore, maintenance inspection of the insulators and selection of maximum insulative values for the insulators is very important.

The *Rolling Sphere Concept* is a protection strategy that is fully described in NFPA 780 and effectively applies to towers up to 150 ft high above level ground. (Figure 2.3-10 shows traditional lightning protection for a studio facility with an adjacent microwave relay tower.) Protective zones are based on an imaginary rolling sphere 150 ft in radius being passed over the structure. The sphere is also rolled around the structure tangent to earth. Air terminals are placed in such a way that the ball never contacts the structure or other objects requiring protection. Properly protected masts, light poles, and adjacent buildings may be taken into account when establishing zones of protection.

The protective zones created by the 150 ft radius sphere have proven statistically adequate for most facilities and are the basis behind most codes and standards.[6,7,8] An interesting way to think about structural lightning protection design is to imagine rolling an inked sphere of 150 ft radius around and over a structure. The ball should roll only on the earth, air terminals and other suitable metallic components which have been connected to the lightning protection system. Any other area which receives ink should be considered unprotected. Where grounded guy wires are present, the sphere rolls over the wire in similar fashion to protect items under the projected zone. This is similar to *overhead ground wire* protection described fully in NFPA 780.

Most broadcast towers, and much of the equipment they support, have sufficient mass and conductivity to resist damage from direct lightning contact. Localized burning, pitting and mechanical forces may occur in the immediate area of contact with the lightning channel; however, the brief duration of a lightning stroke creates little heating in areas away from the point of lightning contact. Figure 2.3-10 shows how air terminals, properly bonded to the tower, may be used to protect more sensitive objects.

Equipment Protection

Control of small potential differences that are deadly to electronic equipment and antenna cable, as measured between active circuitry and grounded media, is key to protecting equipment. Controlling the potential differences to a value below the equipment damage threshold will ensure that the equipment survives. Providing tighter voltage control to a value below the equipment upset threshold will help to ensure that the system rides through the lightning event without any noticeable effect. All protection, however, is predicated on the assumption of average or standard lightning strike parameters such as energy waveform and can therefore, never be 100% effective.

Personnel Protection

Protecting service personnel from the threat of a direct lightning strike, secondary flashing (sideflash), and controlling differences in potential (step and touch potentials) between different parts of their bodies during a lightning event must be integrated into any comprehensive tower protection scheme. Step potentials are voltage gradients seen along the surface or near surface of the earth as lightning current radiates hemispherically from its point of entry into the soil or grounding system. Touch potentials are voltage differences developed in horizontal and vertical elements of a structure, natural object, or system during the passage of lightning current. Both step and touch potentials can be hazardous and must be minimized.

Secondarily, the broadcast tower and associated building should also be reviewed for protection against lightning's effects. In most cases however, towers and equipment buildings will be self protected and provide a means to intercept lightning strikes, conduct the lightning current safely through or around the structure and dissipate the current into the earth. These characteristics however, are critical to personnel and equipment protection. Lightning current passing uncontrolled through a structure may result in deterioration of tower joints, ignition of combustible materials, generation of explosive forces in masonry and other moisture bearing materials and burning or tearing of roofing systems. Secondary flashing between the primary current path and nearby unbonded grounded objects may also pose a threat to persons in or near elements of the structure.

Protection Strategies

A typical broadcast tower installation will consist of an antenna tower in association with a studio building or an equipment hut for repeater locations. The preferred building type would consist of structural steel framing to maximize equipotential grounding and bonding characteristics. Where the buildings are not fully within the zone of protection provided by the tower or other adjacent structure, they must be provided with a Franklin Rod type system. Design requirements for this system are adequately documented and described in NFPA 780, Standard for the *Installation of Lighting Protection Systems*[6] and as supported by UL 96A[7].

Broadcast towers are normally 3 or 4 legged structures constructed of galvanized tubular or structural steel with sectional, vertical truss elements. They are either self-supporting or guyed and can be of various heights from a satellite dish near the ground to a 250 ft. or higher antenna support structure. Lately, however, some architecturally creative type towers of both steel and concrete have been introduced that cause some interesting challenges to lightning protection methodologies. Also, some towers are ground mounted, some are mounted on top of concrete or steel buildings of various heights and some are large enough to span, multi-legged over the broadcast building. It is therefore, impossible to describe detailed protection techniques for all types of towers. Basic requirements, however, will be presented so that application to any type tower can be easily extrapolated.

The tower structure itself is the basic lightning protection element that helps dissipate direct lighting strikes, minimizes voltage rise on associated antennas, coaxial cable and equipment and protects adjacent buildings. Where new towers are located in populated areas, increased lightning activity and dissipation of ground currents can cause local controversy, increased risks and can influence tower design complexity beyond the basic requirements presented herein.

Tower protection elements consist of the following:

- Grounding
- Coaxial cable shield bonding
- Enhanced strike attachment
- Warning light bonding
- Maintenance of joints
- Protection for side and top mounted antenna
- Grounded versus isolated guys

Proper tower grounding is essential to protection of all other systems and facilities interconnected with the tower. It should consist of exothermic weld connections between the tower base or legs and foundation reinforcing steel, anchor bolts, grounding radials, counterpoise rings, guy anchors and adjacent building counterpoise and fences.

Coaxial cable shields from antennas must be bonded to the tower at the top, near the antenna, and at the bottom where the cable leaves the tower. The cable should exit the tower as low as possible to minimize voltage gradient and current dissipation at the bulkhead panel on the building.

A 24 in. air terminal mounted on top of the tower will enhance the probability that the rod will be struck instead of the tower structure or any side mounted antennas, thereby minimizing maintenance from damaged antennas and pitting of tower steel. If the tower is over 150 ft high, side-mounted antennas above this level are vulnerable to direct hits and should be protected by mounting horizontal air terminals above and below the antennas, protruding at least six in. beyond the antenna. Since they are horizontal and located in the end nulls of the antenna pattern, the air terminals should not affect system performance. (see Figure 2.3-9). The warning light(s) on the tower must be

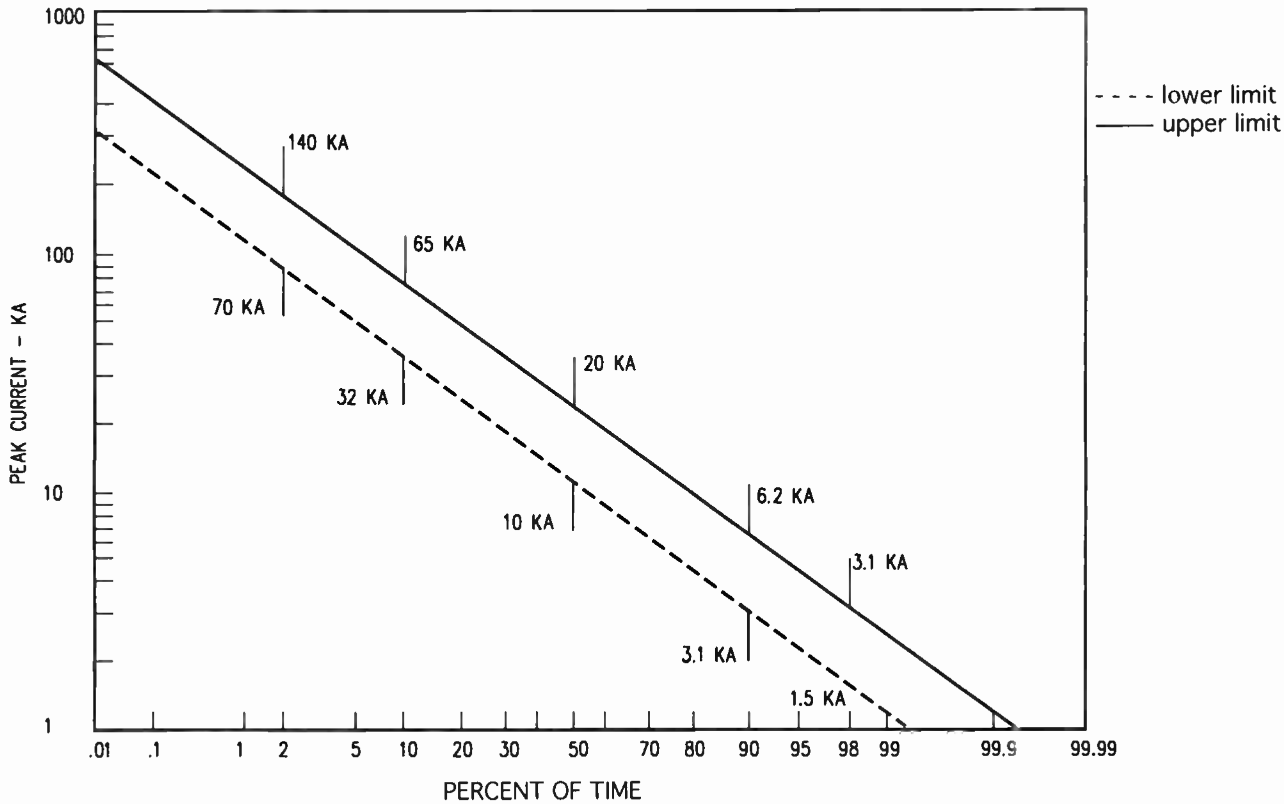


Figure 2.3-6. Lightning Stroke Intensity.

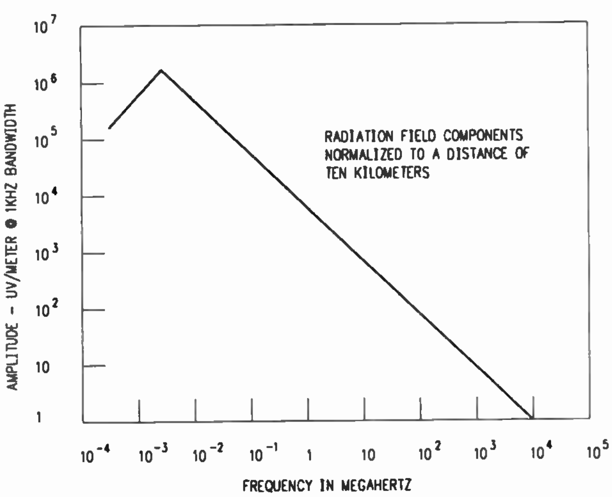


Figure 2.3-7. Lightning signal amplitude versus frequency.

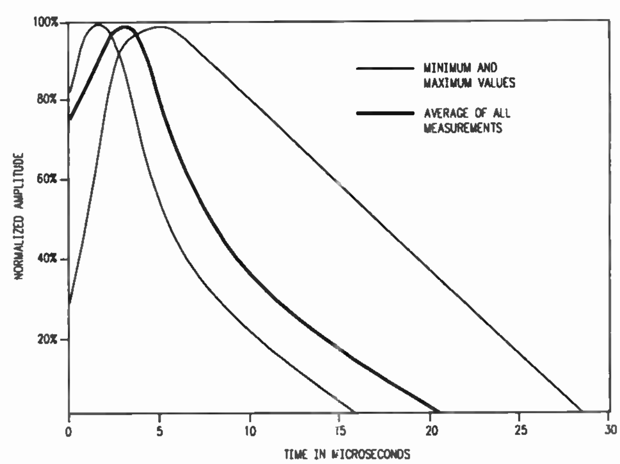


Figure 2.3-8. Typical lightning waveforms.

negative. As in remote strokes, inductive and capacitive properties of the wiring system may cause the circuit to act as a resonant tuned circuit, producing a ringing wave which alternates in polarity. References which categorize the waveforms and current levels for several types of circuits are listed at the end of this chapter.[12–16]

PROTECTION OBJECTIVES

Primary lightning protection objectives specifically for broadcast towers and associated building equipment wiring, may be grouped into two basic, primary categories—equipment protection and personnel protection.

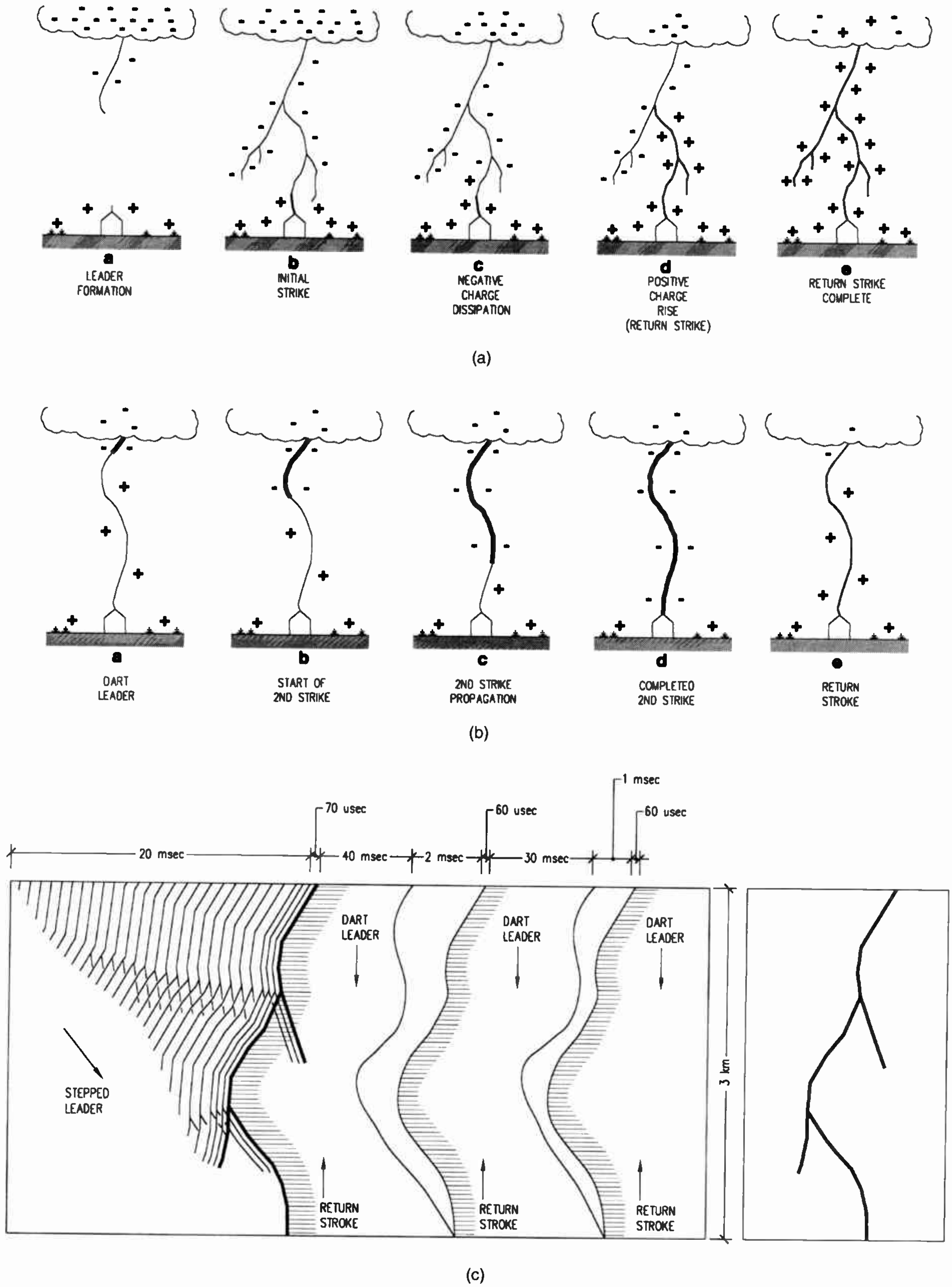


Figure 2.3-5. Typical lightning propagation development: (a and b) step-by-step process of a lightning strike, (c) time development of a stroke and the resulting composite.

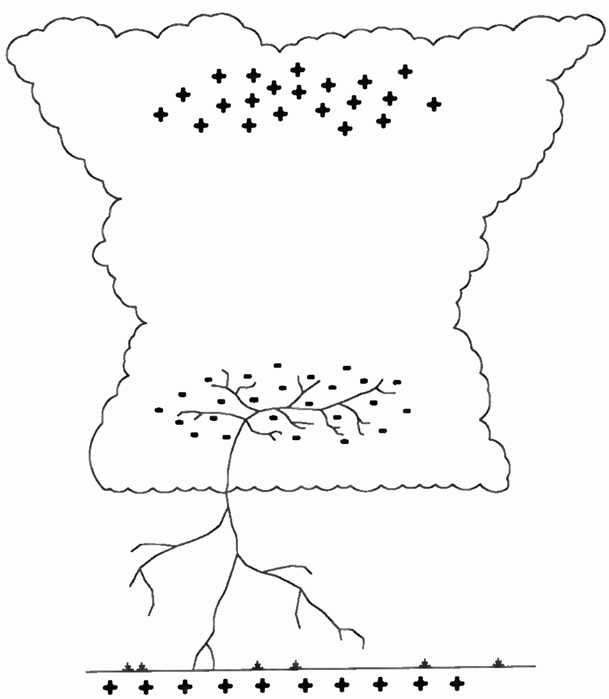


Figure 2.3-3. Typical mode of discharge in summer storms (Negative stroke predominance).

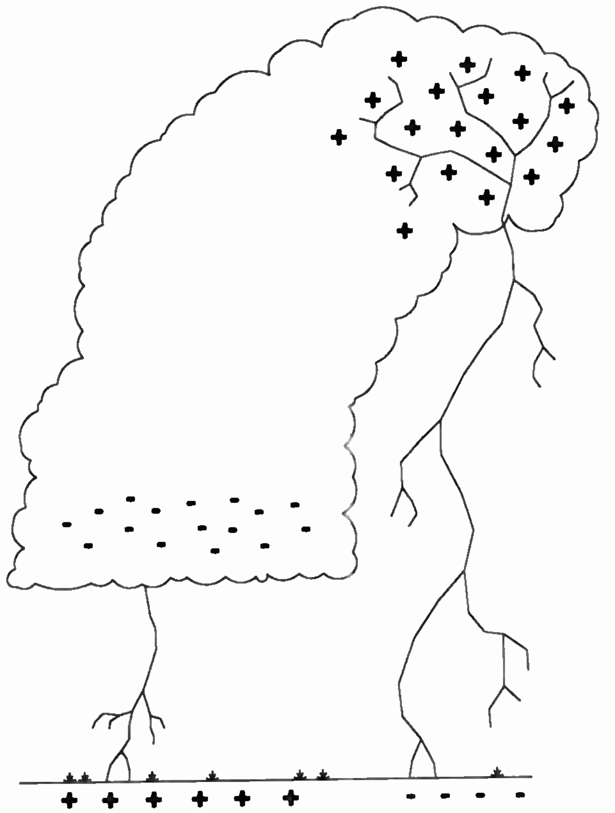


Figure 2.3-4. Typical discharge in winter storms (Positive stroke predominance).

strikes usually attach to the tower structure, building air terminals, tower mounted antennas and tower warning lighting. Remote lightning strikes can inductively couple the surge field into the tower structure, coaxial cable shields, warning lighting conduit systems, building rod systems and antennas. Both types of strikes can cause considerable damage if proper protection is not provided in the design of the tower/building system and in applied surge/lightning protection equipment. Design techniques include proper grounding and bonding, selection of tower/building location, coaxial cable routing and bulkhead design. Protection techniques include proper selection and placement of surge suppression devices for coaxial cables, lighting warning circuit, equipment building power service, tower mounted pre-amp equipment and emergency power equipment as well as air terminal protection for side mounted antennas mounted above 150 ft. and for the equipment building located outside the tower protected zone.

As mentioned above, the radiated impulse field from a remote lightning strike can pose a significant hazard to equipment and systems, particularly those which are interconnected by long lengths of cable. Figure 2.3-7 is a composite of lightning electromagnetic field measurements made by a number of researchers which have been normalized to a distance of 10 km. The figure shows a frequency domain distribution which peaks at about 10 kHz at an intensity of slightly more than 1 V per meter. It is important to realize that nearer strikes can create field strengths many orders of magnitude higher than those shown can. The predominant low frequency component is also very effective in coupling energy into systems of wiring, producing continuous frequency and ringing waveforms due to inductance, capacitance and resonant conditions, even if wiring is buried in the ground.

The time domain current waveform associated with a typical lightning strike is characterized by a very fast leading edge or risetime, followed by a more gradual decay. Technically, risetime is the period of time required for the wave to increase from 10 to 90% of its crest value. Decay time is normally expressed as the time measured between the wave crest and 50% of the crest value. A description of a waveform such as $1.5 \times 50\ \mu s$ for a 10 to 20 kA strike would indicate a single impulse waveform with a risetime of 1.5 μs and a decay time of 50 μs, to half peak value.

There are many variations on actual lightning waveforms seen in real world circuits. Waveforms such as those shown in Figure 2.3-8 may be found with risetimes measured in a fraction of a microsecond near the point of lightning entry to a circuit. Normally, however, the higher the lightning current, the slower the rise time and decay period such that a 250,000 amp strike will have a much slower rise time (30 μs) and 250 μs decay compared to a 10 to 20 kA strike. As a wave propagates through a wiring system, the risetime and the decay time will lengthen. The polarity of the impulse may be either positive or

Figure 2.3-2. 1989–1993 Composite of measured lightning flash density per year. Flashes per square kilometer. (Data supplied by the National Lightning Detection Network.)

field strengths of 100 kV per meter and higher. The field strength above the top of a 50 meter, well grounded, broadcast tower therefore could approach several million volts and result in upward charge streamers and corona discharge. A lightning strike to the tower is then a very real possibility.

The lightning process actually begins as a faintly visible stepped leader originating normally from the base of the storm cell. This leader carries the negative cloud charge toward the positive ground charge, forking and branching due to wind and random charge pockets in the air. As the leader nears the earth, one or more upward streamers will be initiated, usually from the tallest grounded bodies in near proximity to the downward leader. At some point, the downward leader will connect to an upward leader and return stroke currents flow in an effort to equalize the charge difference between the storm cell and the earth. The distance between the connection of the downward leader and the object producing the upward leader is called the *striking distance* and is the basis for the *Rolling Ball* theory. The striking distance is most greatly influenced by the amount of charge in the downward leader such that, the greater the charge, the greater the striking distance. Basically, the amount of leader charge and the surge impedance of the object being struck determine the value of lightning current.

In addition to *bottom of cloud* strikes, severe, long traveling lightning strokes can emanate from the positively charged top of storm cells to negative earth charges. Also grounded objects above 150 ft. tall can initiate upward leading charges, reversing the above process. Grounded objects, such as broadcast towers, taller than 150 ft., can be struck on the side of the tower instead of the top and can subject side-mounted antennas to direct strikes.

Lightning currents range from an average value of 20,000 amperes to a high recorded value of over 400,000 amperes. Many strikes contain up to 30 strokes just milli-seconds apart. These strokes appear to flicker or strobe, are usually wider, brighter and hotter than single strikes and can cause serious damage if not adequately, dissipated. (See Figures 2.3-5(a) through 2.3-5(c) for typical lightning propagation development. Figure 2.3-6 provides probability values of lightning stroke peak values.)

Lightning can affect broadcast tower structures and associated equipment or studio buildings in basically two ways—direct strikes and remote strikes. Direct

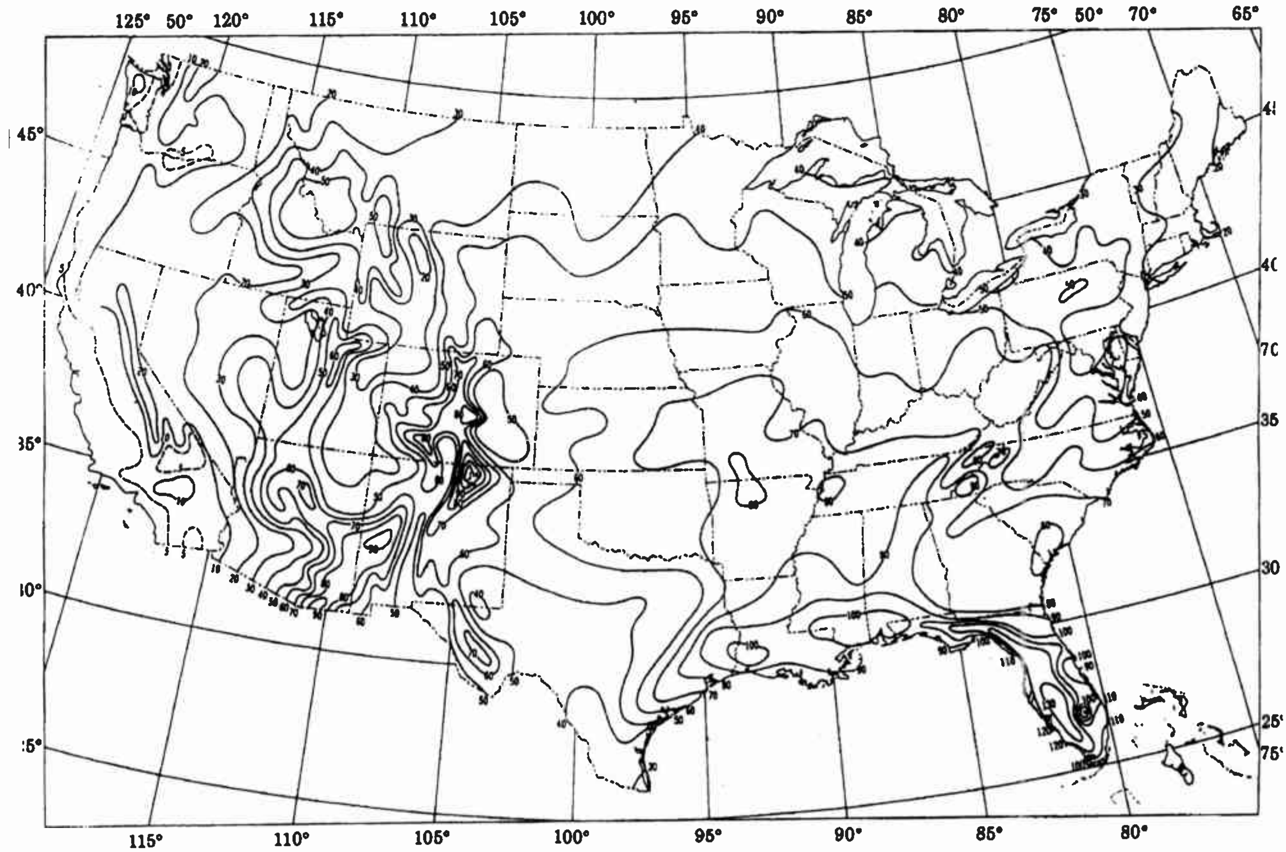


Figure 2.3-1. Average number of thunderstorm days annually. Reprinted with permission from NFPA 780, Installation of Lightning Protection Systems, Copyright 1997, National Fire Protection Association, Quincy, MA 02269. This reprinted material is not the complete and official position of the National Fire Protection Association, on the reference subject which is represented only by the standard in its entirety.

U.S.[1] and other countries have led to new maps depicting ground stroke intensity by location (see Figure 2.3-2). Data is now available in the U.S. and increasingly throughout the world to determine not only whether lightning occurred, but also stroke coordinates, discharge characteristics, flash density, direction of storm movement and other data that can be manipulated into very accurate charts, graphs, trend analysis, histographs, etc. Access to this extensive database can be secured through private and public sources via the Internet and may be useful in selection of broadcast antenna sites and level of protection.

Thunderstorms are of two general types—convection storms and frontal storms.

Convection storms are usually local in extent, of relatively short duration and account for the majority of storms in the U.S. They are caused by local heating of the air near the earth and therefore occur mainly during summer months and in warmer climates where moisture is present, although many convective storms are generated over mountainous areas. These storms can generate to great heights of over 40,000 ft. where vast amounts of moisture and high temperatures are present, producing intense lightning activity. However, they dissipate quickly as accompanying cloud shade and rain cools the earth and dissipates the clouds' source of energy.

Frontal type thunderstorms extend over greater areas, may continue for several hours and are usually more dangerous, causing greater damage than convection type storms. These storms develop from the meeting of a warm, moist weather front with a cold front which may at times extend for several hundred miles, exposing large areas to severe lightning discharges. (See Figures 2.3-3 and 2.3-4 for typical convection (summer) type discharges and frontal (winter) type discharges.)

The formation of lightning in all thunderstorms is generally believed[28] to be caused by ice/dust particles colliding with moisture particles in vertically generated air currents. The collisions create positive and negative charges within the cloud which eventually separate out so that negative charge concentration occurs near the bottom of the cloud and positive charges concentrate at the top. The negative charge at the bottom of the cloud causes a corresponding positive charge on the earth below the cloud. This charge difference between cloud and earth will commonly develop electrostatic

2.3
LIGHTNING PROTECTION FOR TOWER STRUCTURES

EDWARD A. LOBNITZ, P.E.
TILDEN LOBNITZ COOPER CONSULTING ENGINEERS, ORLANDO, FLORIDA

INTRODUCTION

Understanding how lightning works and its effects on tower structures, grounding and antenna systems is very helpful when trying to apply practical protection techniques. *Lightning protection* is never 100% and understanding the limitations or cost effective application of protection is an important part of the design process. Also, since no two installations are exactly alike, repetitive siting of towers will always require a review of *standard* protection to assure that expected protection has not been compromised.

Lightning is potentially one of the most hazardous natural occurrences on our planet. It can kill, maim, start fires, cause explosions, damage equipment, interrupt critical data transmissions, stop a shuttle from flying and cause many other disasters. It is also beautiful, awe inspiring, exciting, mysterious and challenging to all who are enchanted by its technical fascination. It is both predictable and unpredictable. The following statistics are surprising and sobering when we are asked to *protect* against such a power:

From 1959–1996	
Total injuries:	1,337
Fatalities by State:	
Florida:	374
Texas:	169
North Carolina:	168
New York:	128
Tennessee:	126

- A *typical* Florida thunderstorm can unleash 10,000 lightning strikes
- In Florida, lightning kills 10 and injures 35 people every year. It is more deadly, in recent years, than hurricanes, tornadoes or floods
- Nationally, about 90 people a year die from lightning and 300 are injured
- Lightning strikes the earth 100 times each second
- There are about 45,000 thunderstorms annually across the United States
- Property loss is estimated in the hundreds of millions of dollars annually.

Methods of protection from lightning have been changing ever since Benjamin Franklin's original studies of lightning and the creation of his lightning rod system—the Franklin Rod System. Although the exact physics of structure lightning rod protection originally was not fully understood, extensive research has greatly enhanced our understanding of lightning physics while also creating great controversies regarding theory and protection techniques. Also, electronic equipment has entered the picture in one of the fastest growing technologies of our century, but with inherent weaknesses to the effects of lighting. This has created even another huge industry known as surge suppression, with its own mix of technologies, theories, manufacturing methods and jargon to help further the protection methodologies and challenge *protection* designers with almost unlimited technical choices.

Fortunately, there are some standards to help *protection* designers through the maze of technologies available in order to evaluate effectiveness, technology comparisons, test results, performance claims and other criteria. Many of these standards are listed in the references at the end of this chapter.

These standards and publications are considered *consensus* standards since their committee members or testing agencies are comprised of members from across the industry or third party testing agencies. Only equipment that is verified to be in compliance with these standards and testing agencies should be used by protection designers. Claims by *other* equipment manufacturers regarding performance or construction should not be considered unless full verification and proof of claims by third party testing is provided. Lightning can cause great losses to life and property, and assumed protection must be equal to the task without question. More cannot be expected from lightning protection than it can be proven to provide, regardless of guarantee or guarantee claims.

LIGHTNING AND STORM DEVELOPMENT

The intensity of lightning storm activity throughout the world varies by location. Until recently, thunderstorm activity was measured by the number of *thunderstorm days* observed and reported. In the United States, data is reported through the National Weather Service and displayed on an isoceraunic map (see Figure 2.3-1). A thunderstorm day is defined as any day during which thunder is heard at a specific observation point. Since these observations merely confirm the presence of lightning and do not provide information regarding severity, number of ground strokes and stroke location, more complete data was needed to determine exposure risk. Recent atmospheric measurements throughout the

Base Insulator

The porcelain surface should be wiped clean with a soft cloth to remove accumulated dirt. A check should be made for cracks or chips on the porcelain surface. Scratches are often mistaken for cracks. Oil-filled insulators will display a wet surface or leak if cracked. If an oil stain or leak appears at the bottom of the porcelain on an oil-filled insulator and a crack cannot be found, incorrect loading possibly due to settlement of the pier should be suspected. A cracked base insulator should be replaced as soon as practical. Any sign of corrosion in the upper and lower bearing plates, rain shield, or lightning gap should be noted and corrected in a manner similar to that described for the tower structure. The lightning gap should be adjusted in accordance with instructions from the station engineer.

Tower Base and Guy Anchors

The tower base and guy anchors above grade should be visually inspected for spalling and cracking of the concrete. The soil surrounding the tower base foundation should be inspected for evidence of settlement. The anchor arms and surrounding soil should be examined for evidence of movement of the anchor. Any such settlement or movement should be noted.

Steel anchor shafts exposed directly to the soil should be inspected below grade for evidence of galvanic or electrolytic corrosion, especially in areas of high ground conductivity. Extreme caution should be exercised when excavating and backfilling during this inspection to ensure that the anchor's effectiveness is maintained.

Appurtenances

The ladder and its connections should be checked for corrosion and tightness along with the tower. The sleeve and belt of the safety device should be visually examined and tested near the ground level before each use.

Inspection and maintenance of the elevator system should be in accordance with the manufacturer's instructions. It is a good practice to operate the elevator at least once a month.

Inspection and maintenance of the lighting system should be in accordance with the manufacturer's instructions. Checks for corrosion in the conduit, junction boxes, and light fixtures should be made along with the tower inspection. Any obstructions in the breather or drain in the conduit should be removed. Broken or cracked glass and any leaking gaskets should be replaced.

If the tower is equipped with an isolation transformer, its surface should be inspected for cracking and splitting. The surface should be painted with a good quality alkyd varnish. Badly cracked surfaces should be filled with a mixture of varnish and microscopic glass ball powder and the area cotton taped over and varnished.

Frequency of Inspection and Maintenance

A suggested schedule for inspections and maintenance performance is shown in Figure 2.2-10.

A written report of each maintenance and inspection procedure performed should be made and filed with the station engineer.

SUGGESTED INSPECTION AND MAINTENANCE SCHEDULE

ITEM	Daily	Monthly	Before Each Use	Annually	After a major wind or ice storm	Manufacturer's Recommendation
Tower Structure:						
Damaged or deformed members				x	x	
Condition of paint				x		
Corrosion				x		
Connections				x		
Alignment				x	x	
Guys and Insulators:						
Damaged components				x	x	
Corrosion				x		
Connections				x	x	
Tensions				x	x	
					x	
Base Insulators				x		
Tower Base and Guy Anchors				x	x	
Ladder Safety Device			x			
Elevator System						x
Operate		x				
Lighting System						x
Lamp Failure	x					
Conduit Systems, fixtures				x		

Figure 2.2-10. Suggested inspection and maintenance schedule.

Guys and Guy Insulators

Inspection of the guys can be done visually only for those portions adjacent to the anchors and tower. The range of this visual inspection can be extended by using binoculars, but its reliability is limited. If experienced riggers are available, it is possible to ride down the guy on a boson's chair, but this method should be used only under the supervision of qualified personnel.

Other maintenance requirements include the following: A visual inspection should be made of the guy cables, insulators and hardware. Cables and dead end grips should be checked for nicks or cuts in the individual strands. All porcelain insulators should be checked for chips, cracks and oil leaks where appropriate. Fiberglass rods should be checked for surface tracking (black carbon track marks on surface of the rod), breakdown of the epoxy surface and exposure of the individual glass strands. The manufacturers should be consulted with regard to corrective action.

- *Corrosion.* If the guy cables show signs of corrosion, consideration should be given to coating or replacing them. The cost of cleaning and coating the cables should be considered along with the life expectancy of the coating when comparing it to the cost of replacement. All guy hardware should be checked using the same procedures for inspection and corrective actions as previously described for the tower structure.
- *Connections.* All pins should be checked for tightness and the condition of the cotter keys. Dead end grips should be checked to ensure that their ends are completely snapped close, preventing any ice from forming inside. The surface appearance of the guy strand immediately next to the connections should be noted for evidence of slippage. Threads should be given a light petroleum coating.
- *Tensions.* Guy tensions should be checked in conjunction with the tower alignment. These tensions should be measured at the anchor end and compared to the specified values. It is important to remember that they are dependent upon the ambient temperature.
- *Compression* insulators are designed such that the porcelain element is in compression. Simple low voltage types are a single piece of porcelain placed between interlocking loops of the guy. Such insulators are available for mechanical working loads up to 40,000 lbs. For higher loads, oil filled and open types are used. The most common uses of compression insulators in broadcasting are for break-ups and as primary insulators (in-groups of three or four) on low-power antennas.
- *Tension* insulators come in many forms, including porcelain rods (not permitted in structural applications), fiberglass rods, synthetic ropes and oil filled safety core types. Tension insulators are used as primary insulators with corona rings to reduce the electrical field stress at the end fittings. One insulator is required at each guy attachment point. Since the voltage level is different at each point on the tower, different voltage ratings may be required for some insulators at certain guy attachment points. Tension insulators are available in a wide range of electrical and mechanical ratings to meet most needs.

For the usual guy arrangement with cables in three directions, it is necessary to measure the tensions in only one direction while keeping the tower plumb in all directions. For guy arrangements with cables in four or more directions, it is necessary to measure the tensions in only one of the two guys in the same vertical plane while keeping the tower plumb in that plane.

There are several methods of measuring guy tension with varying degrees of accuracy. For small guys up to 3/4 in., a shunt dynamometer calibrated for the size and type of strand is often used.

For larger guys, a series dynamometer may be placed in a temporary line between the anchor and a clamp on the cable. This line is then tightened until the permanent connection is relieved, and the tension is indicated on the dynamometer. Hydraulic jacks with a calibrated pressure gauge or load cells can be used in place of the temporary line and dynamometer. These are particularly effective for large guys attached with bridge sockets.

There are two indirect methods of measuring tensions in guys that do not have any large insulators or other loads in them. The intercept method consists of sighting along a straight bar attached at the bottom of the guy and measuring the vertical distance between the point when the line-of-sight intercepts the tower and the point where the guy is attached. This distance can be accurately estimated by counting the number of bracing panels. The tension in the guy is directly related to this intercept distance, the weight of the guy and its length and slope.

The tension in a guy cable is also directly related to its length, weight and natural frequency of free vibration. The national frequency can be determined by putting the guy in motion with your hand and measuring the fundamental period with a stopwatch. It should be noted that because a guy slopes, the tension on it varies along its length, and this method will only provide the average tension and not the tension at the anchor point. For long cables, this difference can be significant.

All tension measurements should be recorded along with temperature, wind speed and direction. If any substantial changes are noted from the values previously measured, careful checks for slippage of all connections should be made.

Tolerances for guy tensions should be as provided by the designer. In the absence of any other tolerance, tensions should be within plus or minus 5% of the specified values.

Any necessary adjustments in tensions can be made by adjusting the turnbuckle or bridge socket at the anchor. Make such adjustments slowly and carefully. Never leave less than three threads sticking through the turnbuckle body or nut on the socket U-bolt. Remember that the tower must be kept plumb.

isting structural components. The details and specifications for this work must be developed by the structural engineer who analyzes the tower.

Leg members may be strengthened by installing additional bracing or by field-welding additional material to them. Specific procedures in accordance with the American Welding Society *Structural Welding Code* must be provided and followed if field welding is required.

Bracing members and guy cables may be replaced with stronger components. Careful attention must be given to the connections for the new components to ensure their compatibility with the existing tower, as well as providing the required strength for the new components. When replacing components it is essential that temporary bracing or guy cables be installed before removing any existing component, and that it remain in place until the new component has been installed.

It is necessary that damage to the protective finish on existing members due to field welding or reaming be repaired. If required, the affected areas must be painted for aviation obstruction marking.

Foundations and guy anchors are the most difficult components to strengthen, and they may prove to be the limiting factor in determining a tower's capacity. The nature and feasibility of strengthening these components depend on the specific soil conditions.

INSPECTION AND MAINTENANCE PROCEDURES

To ensure trouble-free performance of a tower and its appurtenances, it is desirable to have a regular inspection and maintenance program. Portions of the program can be done by station personnel while others require experienced tower personnel.

Safety precautions should be observed at all times when working on or around the tower. If the tower itself is energized or if a high intensity RF field exists from antennas mounted on it, no work should be done on the tower without clearing it with the station engineer. When climbing the tower, safety belts and climbing devices should always be used. Automatic safety features on elevators should never be bypassed to save time. It is a good idea to never work alone. Failure to observe proper safety measures can result in serious injury or death.

Tower Structure

A visual inspection should be made of the entire tower structure to determine if any of the members have been deformed or damaged. Any bowed or kinked member should be noted as to type, location in tower, and nature and magnitude of deformation or damage. This information should be reported to the tower designer for evaluation and recommended action.

Condition of Paint

A visual inspection should be made of the entire tower structure to determine the condition of the paint. If the painting of the tower is for aircraft observation marking only, and not for corrosion protection, it is necessary only to note any general deterioration rather then small blemishes and scratches. If repainting is necessary, it is important to properly prepare all surfaces and select paints that are compatible with existing finish.

Corrosion

Small scratches in the galvanized surface are not detrimental as the exposed surfaces will be protected by cathodic action of the adjacent zinc. If corrosion is observed, the source should be determined and noted. The affected areas should be wire-brushed clean to bare metal then painted with a zinc-rich prime coat and, if necessary, a finish enamel coat of the appropriate color.

Connections

All bolts should be checked for tightness. Any loose bolts should be tightened in accordance with the original installation instructions.

Alignment

The tower structure should be checked for alignment using an engineer's transit. This check should be done only on a calm day (with wind velocity less than 10 mph), and in conjunction with measuring the guy tensions.

Both plumbness and twist of a tower can be calculated from the measured horizontal deviations of each tower leg member from true vertical. Thus three transit set-ups (one on each leg azimuth) are required for a triangular tower, and four for a square tower. When the transit has been properly leveled, set the vertical cross hair on the edge of the vertical leg at the tower base and lock the instrument in this position. By moving the telescope upward, it is then possible to observe the straightness of the leg over its entire height. The magnitude of misalignment can be accurately estimated by comparison with the tower leg diameter. A record should be made of the observations of each leg at each guy level.

Tolerances for plumbness and straightness should be provided by the designer. ANSI/TIA/EIA-222-F-1996 gives a plumbness tolerance that limits the horizontal distance between the vertical centerlines at any two elevations to 0.25% of the vertical distance between the two elevations. This should never be exceeded. A good rule of thumb in the absence of other data is to keep the tower plumb and straight within the diameter of the leg members. ANSI/TIA/EIA-222-F-1996 gives a twist tolerance of 0.5° in any 10 ft and total twist limit of 5°.

If straightening of the tower is required, it should be done by adjusting the guy wires.

When checking the plumbness of top mounted poles and pylon antennas, the effects of direct sunlight on them must be considered. It is best to make these checks early in the morning or on a cloudy day.

equipment and experienced rigging personnel. It is also important that the firm have adequate insurance coverage including workman's compensation, general and automobile liability, and builder's all-risk for direct damage to the tower and antennas being erected.

Owner's Preparation

Prior to the arrival on site of the erection crew, the site should be made ready for work to begin. These preparations include:

- *Access.* Suitable access from public roads for delivery of the tower materials and erection equipment is required. While a paved roadway is not necessary, the access must be able to handle heavy trucks and construction equipment.
- *Permits.* All necessary building and construction permits should be obtained and posted as required. Any inspections required during construction should be noted.
- *Clearing.* A work area must be cleared to permit unloading, sorting, and assembling the tower. Paths from the tower base to the guy anchors must be cleared for a width adequate to permit hauling the guy cables to the anchors and pulling them to the tower. Paths must also be cleared for the hoist line from the tower base to the hoist location, and for the tag line used to stabilize the loads as they are lifted. The sizes and locations of these cleared areas should be agreed upon beforehand with the erector. A typical layout is shown in Figure 2.2-8.
- *Electrical Power.* Power for operating temporary aircraft warning lights must be available before erection begins.

Assembly

The usual procedure for erecting a guyed tower is to assemble the individual sections on the ground and then lift them one at a time as an assembled unit. For a self-supporting tower, the wider sections near the bottom of the tower are often assembled in the air as the tower is constructed.

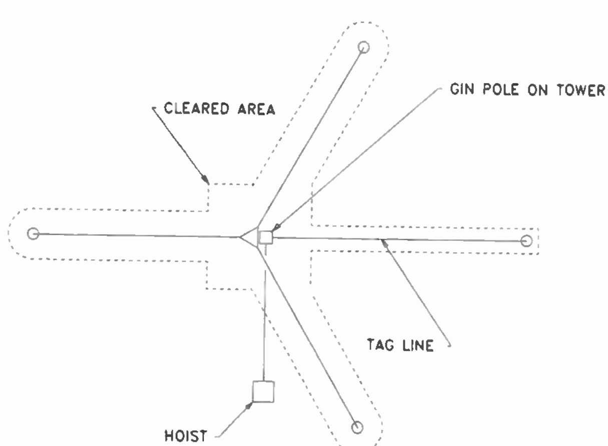


Figure 2.2-8. Typical layout for guyed tower.

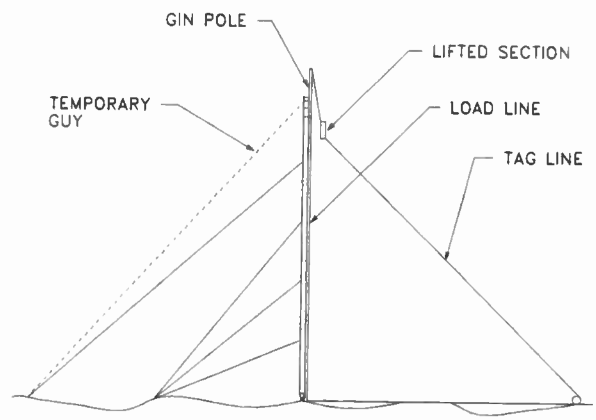


Figure 2.2-9. Typical erection setup for guyed tower.

Assembly of the tower sections should be done on a level bed to ensure that they will be straight and not racked or twisted. Bolts must be properly tightened and have a locking device. For high strength galvanized bolts, tightening by the *turn-of-the-nut* method is preferable to using a calibrated torque wrench.

Stacking

For a guyed tower, the first group of three to six sections are often joined together on the ground and then lifted into place using a crane. This portion of the tower is then guyed with temporary cables, and the remaining section are erected one at a time using a vertical boom or *gin pole*. This boom is moved or *jumped* up the tower as each section is installed. This arrangement is shown in Figure 2.2-9. Temporary guys to stabilize the tower should be used when instructed by the designer.

For a self-supporting tower, a crane is often used to lift as many of the tower sections as possible, after which a gin pole is installed and used for the upper sections beyond the crane's reach.

Temporary aircraft warning lights must be installed at the top of the construction at the end of each day.

The tower should be grounded as soon as the first section is in place.

Guy Installation

When the tower reaches a guy attachment level, the cables at that level are installed. The guys in all three directions should be pulled out simultaneously to prevent any large unbalanced loads on the tower.

The tower should be checked for plumbness as each set of guys is installed and tensioned. Maintaining a plumb tower during erection eliminates the need for time-consuming adjustments later. Final tensioning of the guy cables and a plumbness check are done after the entire tower is erected.

REINFORCEMENT AND MODIFICATIONS OF EXISTING TOWERS

When equipment is replaced, relocated or added to a tower it is often necessary to reinforce or replace ex-ing of the tower, showing where the transmissions lines, conduits, ladder, etc. are located, is also required. The use of presumptive values can result in an analysis with little value.

FOUNDATIONS AND ANCHORS

It is most difficult to predict the cost of the foundation system of a tower installation. This is due to the nonhomogeneous nature of soils, and the uncertainty of the conditions that may exist below grade. Therefore, it is necessary to have an investigation made of the subsurface soil conditions.

It is important to note that the soil design parameters given in the ANSI/TIA/EIA-222-F-1996 Standard are intended to serve only as a basis for preliminary design and estimating of foundation cost prior to obtaining specific soil data. They should not be used for the final design without verification by geotechnical investigation.

Soil Investigation

The soil investigation should be made by an engineering firm, which specializes in soil investigation and evaluation, and is familiar with the general area of the tower site. It should consist of making a test boring at each foundation and guy anchor location, analysis of soil samples taken from the borings, determination of ground water levels, recommendations of parameters for designing the foundations, identification of any special construction procedures required, and recommended backfill specifications. Other soil characteristics or properties may be required because of local conditions, such as the soil resistivity to determine if any special corrosion control methods should be implemented. If piles or rock anchors are necessary, recommendations related to these should be provided. It should also address requirements for frost protection and buoyancy effects.

Because the loads imposed on tower foundations are unique from those for conventional buildings (tower foundations have large uplift and horizontal components), it is important to provide the soil engineer with the loading conditions before they make their investigation. This will enable the engineer to plan their work in a manner suitable for obtaining and reporting the characteristics relevant to designing for the projected foundation loads.

Self-Supporting Tower Foundations

Except for relatively small towers with narrow base spreads, isolated foundations at each leg are usually more economical than a single mat for all legs. These foundations may be spread footings, drilled caissons or driven piles. If sound rock is present at shallow depths it is often economical to anchor the footing to the rock. These anchors should be proof-loaded to ensure their holding capacity in uplift.

Since these foundations are subjected to large uplift forces, it is important to consider buoyancy effects if ground water is present. Also, if driven or cast-in-place piles are used, they must be adequately anchored to the reinforced concrete cap.

Guyed Tower Base Foundations

These foundations may be spread footings, drilled caissons or driven piles. Since they are subject only to downloads with relatively small horizontal forces, they require no special anchorage details for uplift, unless they are placed above expansive soils. Buoyancy is usually not a problem.

Guy Anchors

Deadmen (buried reinforced concrete blocks), drilled caissons or driven piles may be used for these foundations. If sound rock is present at shallow depths it is often economical to anchor the foundation to the rock.

These foundations are subject to large horizontal forces as well as vertical uplift. Therefore, deadmen must have a large enough frontal area bearing against the soil to resist sliding; drilled caissons must have sufficient diameter and depth to prevent excessive lateral deflections as well as pull out from uplift; and driven piles must be sloped to prevent large lateral loads being imposed on them. Rock anchors may be installed along the slope of the resultant of the horizontal and vertical loads, or they may be installed vertically and post-tensioned to clamp the concrete cap to the rock to prevent sliding. Because of the uplift forces, it is important to consider buoyancy due to ground water and to provide adequate anchorage for driven or cast-in-place piles.

Construction

Since nearly the entire foundation system will be below finished grade and not subject to later inspection, it is important to carefully monitor its construction. The following items should be verified:

- Location and alignment of anchors in plan and elevation.
- Condition of excavation surfaces on which concrete will be placed.
- Position, size and grade of reinforcement steel.
- Placement of concrete to prevent voids and air pockets.
- Strength of concrete using test cylinders for 7 and 28 day break tests.
- Protection of concrete against freezing during the curing period.
- Placement and compaction of backfill.
- Driving records and/or load tests of piles.
- Proof loading and post-tensioning of rock anchors.

For towers with extensive foundation systems, it is advisable to retain an independent inspection service for this work. Often the firm making the subsurface soil investigation can also provide this service.

ERECTION

The erection of towers is a highly specialized field and should be performed only by firms having the proper easily controlled for lines installed in the future. Therefore, if the installation location of both the initially installed and future transmission lines can not be regulated, it may be prudent to prohibit the use of this procedure.

Width Restrictions

Some antennas impose restrictions on the width of the supporting tower. One common example is a side mounted FM antenna requiring a maximum width of 18 to 24 in. For antennas with more than eight bays, this results in a very slender structure. When placed at the top of a tall, guyed tower, the design of the guy system for this structure becomes extremely critical. Use of a cantilevered pole structure above the top of the main tower should be considered for these cases.

Another example of width restriction is a panel-type TV or FM antenna mounted on the faces of the tower. Here, too, it is often better to support these antennas on a cantilevered structure above the main tower rather than placing guys within the aperture of the antenna.

As previously mentioned, it is desirable to place large waveguides inside the tower near the vertical centroidal axis to prevent large torsional loads. This requires a tower having a minimum face width in the range of 7 to 8 ft to accommodate the waveguide and its supports.

Initial and Future Considerations

Because the antennas and transmission lines have such a significant effect on the tower design, it is important to consider all possible uses for a tower before it is designed. It is better to have unused capacity than to undergo expensive modifications or replacement in several years to obtain additional height or accommodate another antenna. This has become apparent in recent years with the proliferation of microwave, two-way communications, cellular and personal communication systems.

When providing for multiple antennas, it is important to determine not only the number and type of antennas and lines, but also their location on the tower. The distribution of load is equally important as magnitude.

Triangular top platforms (candelabras) to support broadcast antennas on each corner have been successfully used for many years. They have the advantage of placing all antennas at the same height. A variation of this platform to support only two antennas (tee-bars) has also been used. Both of these systems require multiple guy cables at the top platform to provide adequate torsional stability. It is possible to design the tower for a multiple antenna support platform without installing all antennas at the same time.

Another arrangement of multiple antennas is stacking (installing one antenna atop the tower and arranging others along the tower, one below the other). This arrangement can also be combined with a multiple antenna support platform.

If capacity for microwave antennas is required, it should be provided near guy levels and preferably above to minimize interference with the guy cables. The guy system and web bracing at these levels must be designed to provide adequate torsional rigidity.

Capacity for small antennas may be provided at various locations throughout the height of the tower. One arrangement for a large number of antennas is to provide a platform around the outside of the tower that is large enough to support the radio equipment for these antennas. The antennas can be mounted on the outside railing of the platform, thereby requiring only a short run of coax. Electrical power must be provided to the platform. This arrangement imposes a large concentrated load at the platform location with a relatively small uniform load between the tower base and the platform. If the same number of antennas were mounted along the tower and each fed by an individual coax line from the base, there would be only small concentrated loads at the antenna locations, but a relatively large uniform load due to the lines. This is an entirely different distribution of load, and would have a pronounced effect on the design.

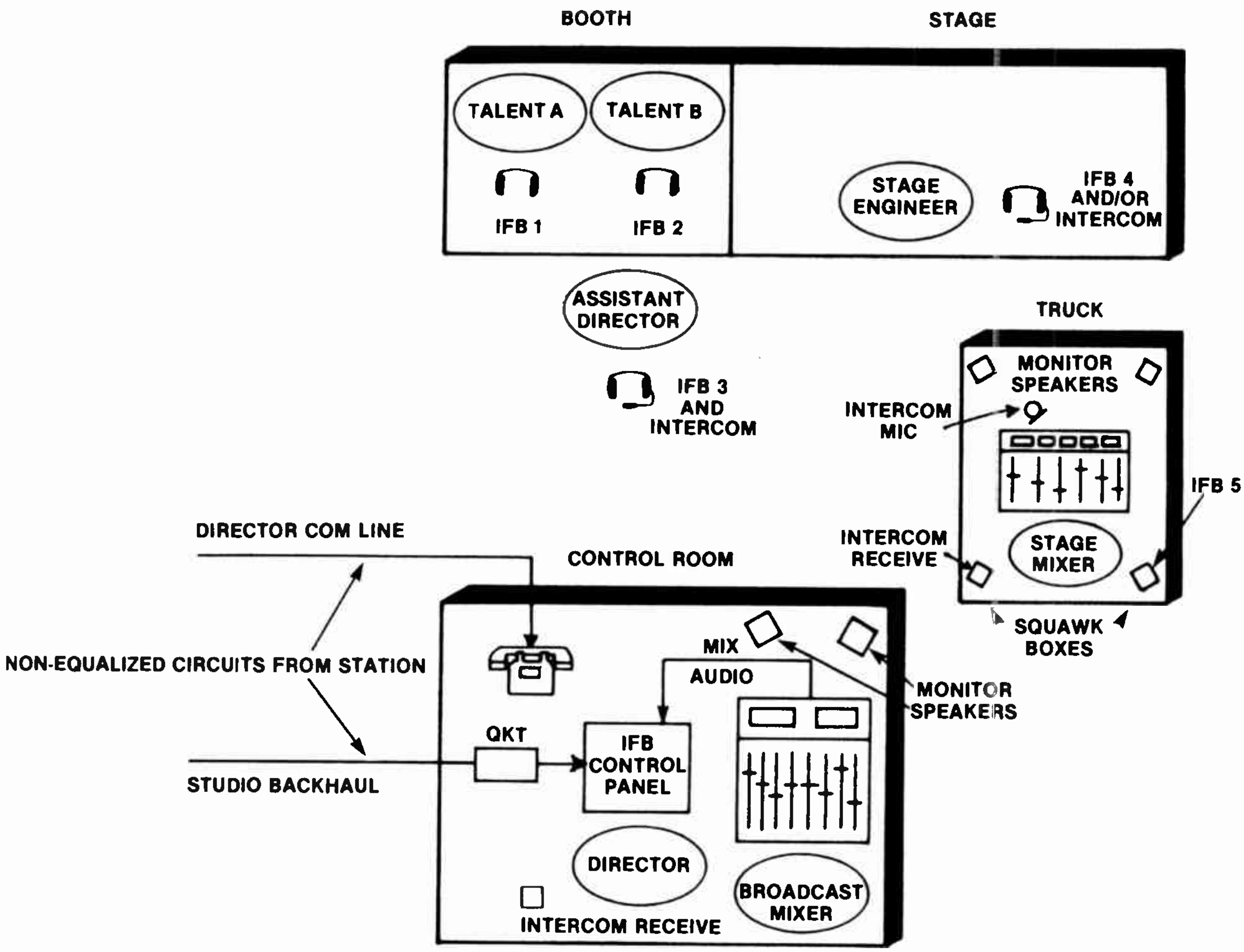
Another important consideration for future antennas or height extension is the electrical system. If an extension in height is planned, the wiring for the aircraft warning light system should be designed so that any additional lights can be connected to the system without adding or replacing wires in the existing conduit. The same holds true for any circuits required for future antennas. If the necessary wiring cannot be provided during the initial installations, capacity should be provided for additional conduits to hold the future circuits.

Replacement, Relocation, or Additions to Existing Towers

Since every tower has been designed for a specified arrangement of equipment, changes should not be made without considering their effects on the structural adequacy of the tower.

Two common misconceptions related to changes in equipment are "lower is better," and "smaller is better." Neither is necessarily correct, especially for guyed towers. Decisions based on these premises can have serious consequences.

It is much better to have a structural analysis of the tower made by a structural engineer experienced in tower design. Because of the significant changes that have been made in the methods of specifying loads in the various revisions of the design standard, the analysis should be made using the same criteria for wind and ice loads used for the original design and also for the current revision of the standard. This analysis will determine if any overstresses would occur in the tower or its foundations, and what modifications and reinforcing would be required to retain the structural integrity. To perform this analysis, it is necessary for the engineer to have complete data on the tower and its foundation including configuration, member sizes, and material strengths. It is also necessary for the customer to provide the engineer with a complete inventory of existing antennas and transmission lines on the tower along with the proposed equipment. A plan view draw-



NOTE: DIRECTOR'S CUES TO BROADCAST MIXER ARE ACOUSTICAL BECAUSE THEY ARE IN THE SAME ROOM AND NEITHER IS WEARING HEADPHONES. THE DIRECTOR CAN TALK TO ANY OF THE FIVE IFB STATIONS INDEPENDENTLY OR COLLECTIVELY. PROGRAM AUDIO SELECTION CAN BE VARIED BETWEEN EACH IFB STATION. ALL INTERCOM MICS ARE PUSH-TO-TALK.

Figure 3.9-15. Communications plot for a complex live remote broadcast. Stage mixer can use simple intercom (with no program audio) for communications to stage and other personnel, while main control room uses interruptible foldback system (IFB) to coordinate entire broadcast.

If critical, multi-station monitoring and communication is necessary at the remote site, an IFB system may be required. This device combines the function of intercom and monitor, such that a director, either on site or elsewhere, can communicate with and cue various personnel at the remote (talent, stage managers, floor directors, engineers) via headset or loudspeaker, using a multi-station intercom. The director's control panel for the IFB system is typically equipped with push-to-talk individual station and *all-call* (talk to all stations) buttons. What differentiates the IFB from a standard intercom is that when the director is not talking to an IFB station, that station receives program audio. The director's cues temporarily interrupt the program audio, which returns after the director's message has ended, hence the term IFB. Unlike a traditional intercom, the IFB system is usually not bidirectional, but rather feeds one-way from the director out to the various receive stations. Figure 3.9-15 shows a typical IFB layout in a complex remote situation.

Most IFB systems allow the use of two or more different program audio sources to be selectively routed to different stations, so that one station could be fed the dry stage mix, for example, while another station heard the whole remote transmission with continuity included. Both would hear the director when their channels were designated for communications, regardless of which program audio channel were selected. Off site IFB (in which the director is not at the remote location but back at the studio) is often referred to as a (PL) *private line*; dial-up or leased telco lines are used for PLs, along with wireless return links or cellular phones, in many ENG cases.

Some IFBs are stereo capable. Besides allowing the system to provide a stereo program feed to all stations, these systems are usually set up to place the director's communications into one channel only. This makes it easier for talent (listening on stereo headphones) to distinguish between program and communications audio. It can also allow the communications to interrupt only one channel, keeping program audio continuous (or only slightly attenuated, if desired) in the other channel.

Set-up and Testing Procedures

A tone oscillator is useful in setting up the remote broadcast. Most mixers designed for remote applications include an oscillator that can be switched on to the program channel. This feature is especially useful when a telco line is employed to return program audio to the studio. If the oscillator has multiple frequencies, so much the better. Phone line tests for frequency response, S/N, headroom, and relative polarity for stereo broadcasts should be performed. Phase response and distortion tests are also useful, if time and test equipment permit. (See Chapter 3.11, *Common Carrier Audio Program Services*, for detailed testing procedures.) If any noise reduction or other enhancement devices are being used across the line, testing should be performed with them bypassed. After the line proves satisfactory, engage these devices and recheck.

Multi-frequency checks are less critical with digital backhaul systems, because these are typically quite flat within the passband. Nevertheless, simple level and polarity checks are still worth performing with ISDN and T-1 services.

Having the proper monitoring facilities is important to the success of any remote broadcast. A loudspeaker (or well-matched pair for stereo broadcasts) and a set of headphones should be provided for the remote crew. Not all portable mixers can support a loudspeaker and multiple headphone outputs, and so a separate power amplifier and headphone booster may be needed. Speakers and headphones selected should be reliable and familiar. Crew members should listen to familiar music over these while setting up, to attune themselves to the speaker in this environment, however, since different rooms and speaker placements will produce drastic changes in the sound of any speaker.

After checking phone lines or wireless links for continuity back to the station, setup the monitoring system first, and check it for clean audio. Then add the other elements or subsystems of the setup one at a time, checking the monitors for continued clean response after each. In this way, when any deleterious effects are heard (hum, buzz, hiss, etc.), it will be fairly simple and quick to track down the offending hardware or interface method.

An off-air receiver is a requirement for nearly all live remote broadcasts. The receiver gives the remote crew a way of checking the total link and allows easy cueing of talent at the event.

A separate dedicated telephone set is suggested for complicated remote broadcasts. The phone provides an easy means of communicating with the studio. It can also serve as a backup line for program audio in case the RPU system or telco program service should fail. And in cases where the remote broadcast originates from outside the coverage area of the station, it is essential for monitoring as well as communicating. If digital telco service is employed, it is generally not much more expensive to have this return (or *backfeed*) line be of the same fidelity as the transmit circuit.

Wireless Microphones

The use of wireless microphones to free up the talent at a remote broadcast is gaining popularity with stations involved in RENG activity. The advantages to the talent are obvious: complete freedom of movement and nothing to carry around but a microphone and air monitor receiver. There are no controls or meters for talent to worry about. The range of a wireless mic is somewhat limited, but a properly designed system for remotes that are more or less stationary can provide simple setup and coverage of an event.

The receiver used in conjunction with the wireless microphone may use either diversity or non-diversity reception techniques. A non-diversity receiver is used where multipath cancellation is not a problem, such as in open areas or when conducting fixed position interviews. If, on the other hand, the wireless mic is to be used in several places and the possibility of multipath cancellation exists due to nearby reflective objects, a diversity receiver is recommended.

The diversity receiver uses two antennas that are physically separated by a distance varying between several inches to several dozen feet. In some diversity receivers, two complete RF sections are used. The diversity receiver automatically selects one of the two signals based on its criteria for better reception. (Some receivers combine the two signals through a phase shifting network.) The switching of RF sources occurs silently without any *squelch type* noise bursts.

Many wireless microphone systems include audio companding circuits to extend the dynamic range and lower the apparent noise floor. A properly engineered wireless microphone system can be treated by engineering personnel as essentially a piece of wire between the microphone and the audio console input. Both VHF and UHF frequencies are used. An often overlooked FCC regulation requires licensing of most wireless microphone systems.

EMERGING TECHNOLOGIES

As always, new technical developments will find their way into the radio remote kit. A number of these systems are emerging at this writing that will likely affect both backhaul and remote production systems.

Backhaul Systems

Under Part 15 of its Rules, the FCC has established two bands for unlicensed use of spread-spectrum transmission (§ 15.247), one in the VHF band (902–928 MHz) and one in the S-band (2400–2485 MHz). Some digital audio transmission equipment has become available for these bands, particularly the latter, which allows RPU-like operation without licensing or frequency coordination. With Yagi type antennas, these low power (<100 mW) systems can be used on line-of-sight paths of approximately 20 miles. Their use is not yet widespread in broadcasting, but other industries (primarily wireless computing) are beginning to use these bands. While spread-spectrum transmission is fairly resistant to interference, there is some concern that as the bands become more popular, the reliability of such audio links will not remain adequate for broadcast use. Nevertheless, in most cases, the audio quality and robustness of S-band spread-spectrum RPU has proven more than adequate for broadcaster's requirements, among the relatively few using the systems to date.

Another emerging telco technology is called *digital subscriber line* (DSL) technology, which extends the capacity of standard twisted pair circuits to T-1 data rates and higher. DSL is being implemented in a number of different forms. The most common to date is the *asymmetrical digital subscriber line* (ADSL), which provides high-speed service in one direction and lower speed service in both directions, plus bidirectional POTS service—all on standard copper telco loops (two twisted pairs). ADSL can provide speeds ranging from 1.5 Mb/s downstream and 64 kb/s upstream (plus POTS) up to 8 Mb/s downstream and 1 Mb/s upstream. *High-bit-rate DSL* (HDSL) is a symmetric service, offering from to 2 Mb/s to 6 Mb/s in both directions. *Very high-bit-rate DSL* (VDSL) offers extremely high data rates over shorter distances (13 Mb/s up to one mile, or up to 51 Mb/s at 1,000 ft.). DSL (particularly its HDSL variant) could have significant value for wired remote audio backhaul without data compression. (For more detail on DSL, see Chapter 3.11, *Common Carrier Audio Program Services*.)

Remote Production Equipment

In the area of location audio hardware, one new system that may have application is the *fiber optic snake*. Here, a single, robust, yet lightweight cable carries several dozen separate microphone or line level channels from one part of the remote site (typically the stage area) to another (typically the mix position), replacing the traditional heavy copper multipair cable. Also unlike its predecessor, the fiber optic system is impervious to the pickup of EMI, nor does it generate any. Microphone level audio is preamplified at the head or stage end, then digitally encoded and multiplexed, converted to the optical domain, and sent down the fiber. The reverse process takes place at the console end, terminating in line level analog balanced outputs, to be plugged into a conventional mixer's line inputs. Splitting can also be accomplished without difficulty or degradation in the digital (electrical, not optical) domain with these systems. A helpful feature is the addition of some remote control ability for microphone preamp gain from the mix position, since the preamps are located at the head end of the snake. Because this is an active rather than passive system, some provision must be built in for return signals running in reverse from the console back to the stage. A typical configuration is 56 send and 8 return paths on a single cable, about the size of a standard mic cable.

Also new and of potential future value is the digital wireless microphone. Using unlicensed S-band frequencies, these systems can improve the audio quality and coverage of wireless mics used by broadcast talent. Similar advantages may be applied to future digital wireless IFB systems.

CONCLUSION

A RENG network should be planned and constructed with long term service and frequency coordination requirements in mind. Areas that currently do not experience spectrum congestion problems may encounter them in the near future. It pays, therefore, to design a system that is spectrum efficient and relatively immune to interfering signals. New digital services, both telco supplied and of an RF variety, should also be considered, and the progress in these areas carefully monitored for application on future links. It is always easier—and cheaper—to do the job right the first time.

No matter how good the backhaul link, if the remote audio program is not properly created at the site, it will not make a successful broadcast. Well equipped, simple and reliable systems are essential for the remote crew to be able to setup quickly and make the instant creative decisions often required on location. The well engineered remote can make radio programming exciting and unique—the kind of compelling material that keeps listeners coming back for more.

BIBLIOGRAPHY

"Common Carrier Audio Program Services," *NAB Engineering Handbook*, 9th Edition, Chapter 3.11.

Pizzi, Skip., "Beyond ISDN," *BE Radio*, January 1998.

Pizzi, Skip., "Dealing with Telco Lines (Parts 1 and 2)," *Mix*, April/May 1985.

Pizzi, Skip., "Engineering Radio Remotes," *Broadcast Engineering*, January 1986.

Pizzi, Skip., "Remotes Revisited," *Broadcast Engineering*, January 1991.

3.10
TELEPHONE NETWORK INTERFACING

STEVE CHURCH
TELOS SYSTEMS, CLEVELAND, OH

INTRODUCTION

From earnest political talk presentations to raucous morning shows, listener involvement via telephone is an important programming element at many radio and television stations. When we want to create a two-way connection with our listeners, we will probably be using the dial-up telephone network.

Radio news departments rely extensively upon phoners to get reporters and newsmakers on the air in a timely fashion. Why are the people who run local TV news so concerned with avoiding the dreaded talking head—that is, the anchor simply reading a story into the camera? Because they've discovered that *being there* is better. The same is true for radio.

Today, integrated services digital network (ISDN) lines combined with modern audio compression techniques permit instant full fidelity remotes from almost anywhere in the world.

This chapter will explore all of the ways to integrate the ubiquitous telephone network into broadcast operations. First, we'll learn about the nature of the various services available from telephone companies. Then we'll investigate ways to interface them to our station facilities.

THE TELEPHONE NETWORK

As we transition our broadcast facilities to digital systems, it is interesting to note that the standard voice telephone network is almost entirely digitized and has been so for many years. The watershed event was Illinois Bell's 1962 installation of a T-carrier system—the first widespread commercial application of digital audio. Telephone engineers appreciate digital technology for the same reason broadcasters do: reduced susceptibility to noise and other disturbances, and improved ability to switch, monitor and maintain the circuits.

While the worldwide dial-up telephone network is an amazing achievement, it is mostly made from a simple ubiquitous element: digital circuit-switched channels of 64 kbps each. *Circuit-switched* means that the channel is connected end-to-end with the entire capacity available for the duration of the call. (This is in contrast to packet-switched systems, such as the Internet, where capacity is shared among users and there is usually no guaranteed bandwidth.)

While most of the network infrastructure is digital, the last-mile copper connections from the central office to the customer site mostly are not. The vast majority of users interface to the network via an analog technology that is little different from that employed in Alexander Bell's days. This is beginning to change with the introduction of digital last-mile technologies like ISDN, T-1, and an Asynchronous Digital Subscriber Line (ADSL).

Incidentally, in industry jargon, your local phone company is a local exchange carrier (LEC) or simply a telco. A long distance company is an inter-exchange carrier (IEC).

Speech Coding

The bit rate of 64 kbps was chosen to support phone-grade speech audio encoded using a modified pulse code modulation (PCM) technique. When we make a plain old telephone (POTS) call, our speech is sampled at an 8 kHz rate and encoded into a digital word 8 bits long. Telco engineers call this 64 kpbs bitstream a digital signal level 0 (DS-0) channel.

The word length is what determines dynamic range—and 8 bits would only permit 48 dB were it used in standard PCM linear fashion. A primitive kind of compression is used to stretch the dynamic range: μLaw in North America and much of Asia, and A-law in Europe (see Figure 3.10-1). This is a scheme that equalizes the step size in dB terms across the dynamic range—a smaller step size on low level signals reduces quantization noise and improves effective dynamic range to the equivalent of about 13 bits. Thus, the quantization noise (and distortion) is approximately a fixed percentage of the signal amplitude, regardless of its level.

The process of conversion and companding is done in specialized analog-to-digital (A/D) and digital-to-

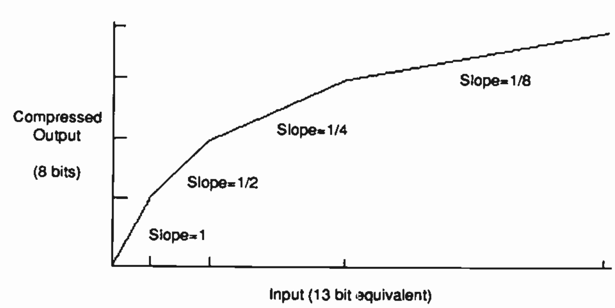


Figure 3.10-1. μLaw PCM coding within the telephone network causes the noise to be approximately a fixed percentage regardless of level.

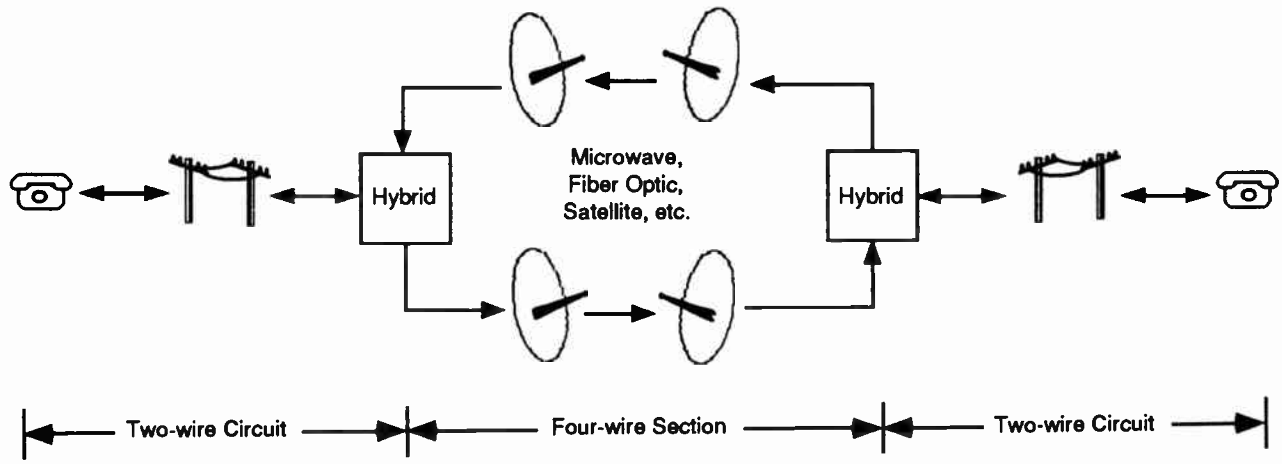


Figure 3.10-2. Two-wire circuits have both directions on a single pair of wires, which are separated for switching and long-distance transmission into 4-wire signals with hybrids.

analog (D/A) integrated circuits called *codecs* (CODer/DECoders). The method is specified by the International Telecommunications Union (ITU) as standard G.711.

2-Wire and 4-Wire

Both speech directions are mixed together on the usual analog lines with which we are most familiar, but this is not the way signals are handled within the telephone transmission and switching network. Non-copper transmission media such as microwave radio, satellite and fiber-optic cables are one-way only, so the paths must be kept independent. Even when copper is used, long-distance links are kept separated so that amplification can be inserted. A standard analog POTS circuit is *2-wire,* because it arrives on two wires. The network is internally *4-wire,* so named because in the past, a 4-wire circuit needed a separate wire pair for each of the send and receive transmission directions—four wires altogether.

The Traditional Analog Line

The traditional telephone lines provided by the phone company are known officially as subscriber loops, trunks or simply CO (central office) lines. (Trunks used to refer only to lines destined for private branch exchange (PBX) systems and may have included special signaling as well.)

Because these are 2-wire circuits, the CO uses a 2-to-4-wire converter (also called a *hybrid*) to interface the analog lines to its internal 4-wire system, as shown in Figure 3.10-2. This process happens on the *line card,* which is also responsible for digitization, talk battery insertion, off-hook detection, and ring generation.

Talk Battery and Ringing

The *talk battery* direct current (dc) voltage and the conversation audio appear together on the phone pair. The talk battery leaves the exchange at −48 V and is limited to 20–50 mA by a series resistor. The resistor's value is selected to complement the resistance of the loop. The dc resistance of the loop itself varies from a few to 1,300 Ω depending on length. Because of this series resistance, when a line is off-hook, its voltage at the customer equipment drops to around −12 V.

For ringing, an ac voltage of 90 vrms at 20 Hz is superimposed on the line. Talk battery is maintained during ringing, so that the resulting signal has a sinusoidal shape shifted 48 V to the negative.

Talk signals are ac coupled with nominal impedance of 600 Ω. However, some CO equipment uses complex impedance coupling, and the nature of the telephone network usually results in the actual impedance as presented to the user rarely being the specified simple 600 Ω. This turns out to be an important issue for broadcast interfacing, which we will discuss in detail later. The basic parameters are summarized in Table 3.10-1.

Frequency Response

For ordinary subscriber loops, the phone company specifies a frequency response of 300 Hz to 3.4 kHz. In the not-too-distant past when all local calls were connected at the exchange by metallic contacts, better

Table 3.10-1
Phone loop characteristics.

Parameter	Typical U.S. Values	Operating Limits
Talk Battery Voltage	−48 VDC	−47 to −105 VDC
Loop Current	20 to 80 mA	20 to 120 mA
Loop Resistance	O to 1300 ohms	0 to 3600 ohms
Loop Loss	8 dB	17 dB
Distortion	−50 dB	N.A.
Ringing Signal	20 Hz, 90 VRMS	16 to 60 Hz, 40 to 130 VRMS
Noise (objective)	−69 dBm0 to 180 mi, −50 dBm0 to 3000 mi (−16 dBm0 talk level) (C msg weight)	

frequency response was likely to be had on many conversations. Today almost all calls are digitized and are strictly limited to a 3.4 kHz bandwidth by the sharp low-pass filters required for proper digitization. The phone network's 8 kHz sampling rate permits a theoretical Nyquist frequency of 4 kHz, but a 600 Hz transition band is necessary for anti-aliasing and reconstruction filtering (see Figure 3.10-3).

Noise and Level

A 1971 Bell System survey of the phone network nationwide determined that the average conversation had a level of −16 dBm. Of course, as anyone who has wrestled with broadcast-to-telco interfacing knows, incoming level varies tremendously, with a range of perhaps −40 to −4 dBm, as illustrated in Figure 3.10-4.

Send audio (that is, audio fed into the telephone line) must be limited to −9 dBm as specified in Part 68 of the Federal Communication Commission (FCC) Rules. Audio loss on any given local loop is limited by tariff to 8 dB or less. This loss limit, however, applies only to the loop from the CO to the subscriber and does not include the rest of the signal path. Also, the 8 dB loss may occur at each end of a conversation path: once at the calling party end and again at the called party end, for a total loss of 16 dB.

The phone engineering people measure noise upside-down, defining a reference noise floor and then measuring *up* from there. The reference noise level is one picowatt, which corresponds to −90 dBm. Thus, a noise level of −60 dB relative to 0 dBm would be reported as 30 dBrn noise (dBrn = dB above reference noise). Note that, according to this method, the higher this number, the worse the noise.

Be aware also that when telephone people measure noise, they are measuring only *idle channel noise*. This is an important difference, since in digital systems idle channel noise is not the same as the traditional (S/N) measurement in analog systems. Noise in a digital system will generally increase when a signal is present. This effect is called *modulation* or *quantization* noise and is primarily dependent upon the number of bits used for quantization.

A *C-message weight* filter is employed when measuring phone line signal-to-noise ratio (S/N). (See Figure 3.10-5.) The C-message curve was developed years ago to simulate the frequency response of an old-style telephone earpiece and, accordingly, it has considerable low-frequency roll-off. This means that a line can have significant hum and other low frequency noise and can still meet the officially mandated noise specs. While this makes life easier for the phone company technicians, it can be troublesome when a broadcaster is trying to use phone audio on the air. If noise is a serious problem, try to get the technician to switch the noise meter to the flat position. The measuring set usually does have this option available.

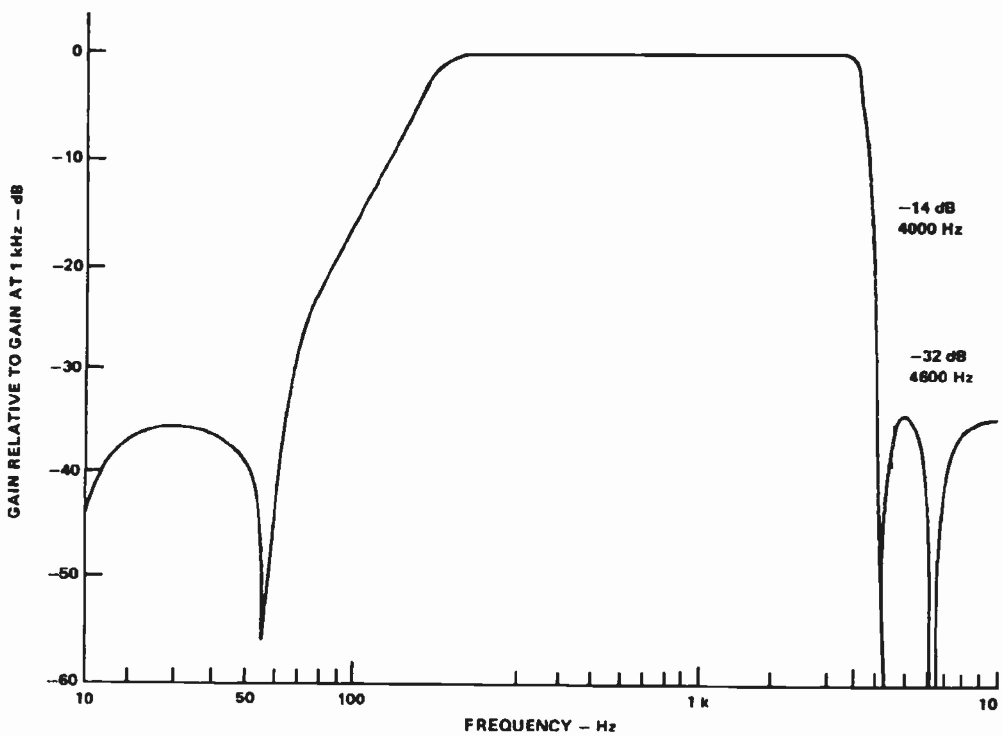


Figure 3.10-3. The low-pass filters required for digital transmission restrict frequency response. This response curve is for a codec that is widely used in the telephone network. (Note also the significant low frequency roll-off).

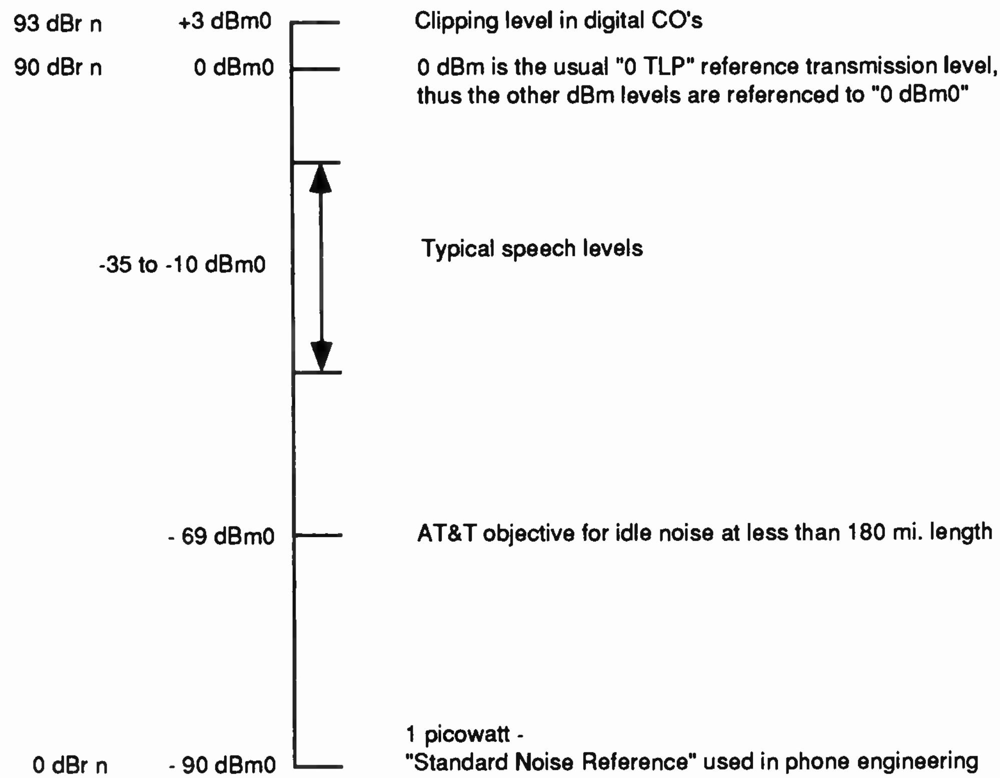


Figure 3.10-4. Signal and noise level references used in telephone engineering.

DTMF Tone Dialing

Dual Tone Multiple Frequency (DTMF) dialing uses two frequencies for each digit in order to avoid talk-off—that is, the tone detector accidentally sensing voice as a dial command. In addition, the frequencies were carefully chosen to avoid problems with harmonic distortion causing false detection. There are four low group frequencies, one for each button row, and four high group frequencies, with one assigned to each column as shown in Figure 3.10-6. Tolerance is ±1.5% for the encoder and ±2% for the digit receiver. The time required to recognize any digit tone is 50 msec with an interdigit interval of another 50 ms. Low group tones are supposed to be sent at a level between −10

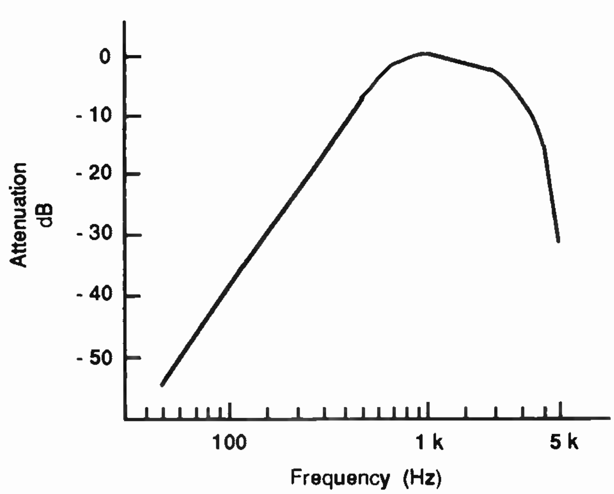


Figure 3.10-5. C-message weight frequency response curve.

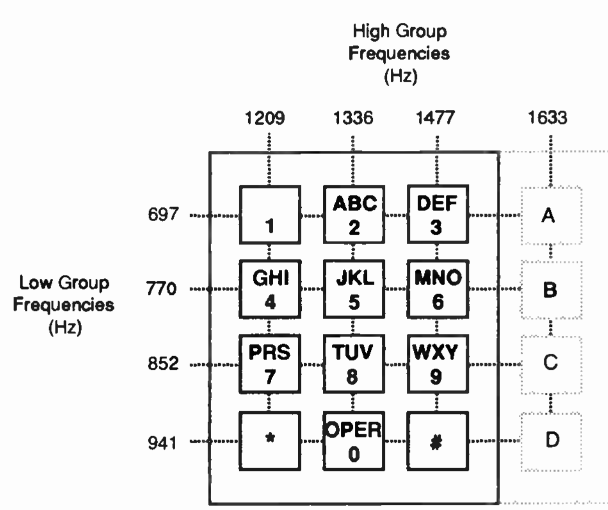


Figure 3.10-6. DTMF tone keypad frequency assignment. The four tone pairs in the last column are for special applications.

and −6 dBm; ideally, tones in the high group are transmitted with 2 dB greater level in order to compensate for high-frequency roll-off in the phone line.

Loop Start and Ground Start

Central office lines come in two basic configurations: loop start and ground start. Loop start is the kind that is most common. In this kind of circuit, the CO provides talk battery to the line at all times and detects that an off-hook condition is occurring when the terminal equipment connects and causes current to flow between the tip and ring. (Incidentally, the terms *tip* and *ring* originated with the description of the circuits being on the tip and ring of the patch cords that used to be used by telephone operators.) With ground start circuits, the CO waits for a connection from the ring wire to ground before connecting talk battery, at which time the terminal equipment removes the ground connection to establish a balanced talk path. When the calling party hangs up, a ground start circuit removes talk battery. A loop start circuit may or may not provide a momentary interruption or reversal of the talk battery when the calling party terminates.

Many PBXs are designed to work with the ground start circuits because the possibility of collision is reduced. Collision occurs when the phone system tries to seize a line for an outgoing call just as that line is ringing in.

Disconnection: Calling Party Control

Loop-current interruption occurs on most telco lines when the calling party hangs up. It is sometimes referred to as calling party control (CPC), since the *calling party* controls *your* equipment when he hangs up. The CPC may turn off an answering machine, for example, or extinguish the winking light on a held line on a key phone. The CPC interruption was probably never intentional, having been a by-product of early mechanically switched relay-controlled exchanges. Thus, some phone lines do not provide this function or they provide it unreliably. However, with the proliferation of answering machines that rely upon CPC, most central office equipment now has this capability designed in. In some cases, it is necessary to specifically request this feature from the phone company on a per line basis.

Loop-current *reversal,* on the other hand, has long been a phone company signaling method. First used between the telco's own central offices, loop-reversal was later employed to communicate with some large premises PBX systems. Thus, lines that are set up for PBX use, or originate at central offices with large concentrations of business customers, sometimes use this method. (However, the preferred and more modern situation for PBX control is to use ground-start lines.)

While most exchanges do provide CPC, there are some that do not reliably provide it or provide it after a variable time delay. Most PBXs do not generate it. However, every telco CO in the United States eventually returns dial tone to its lines when the calling party hangs up. Thus, we can use the presence of dial tone as a back up to cause a disconnect when the loop-current detection methods fail. An important consideration is to prevent false talk-off from noise, applause or other spectrally rich audio. Using software based statistical methods ensures that the dial tone is *really* present before terminating the connection.

Caller ID

Caller ID (CID) allows you to know the phone number of the caller. This capability is useful for call-in shows, where it might be desirable to deny access to problem callers. The technology is simple. Between the first and second ring, the information is sent in a packet using a 1200-baud modem. This is exactly the same modulation scheme used in normal computer modems operating at this rate. Customer equipment normally suppresses the first ring so that the answering user does not take the call before the CID information is fully transmitted.

Loading Coils

A typical #24 gauge phone pair attenuates a 3 kHz signal 2.5 dB per mile due to capacitive effects. On an 8 mile (12.9 km) long line, high-frequency attenuation would thus be 20 dB, a significant amplitude distortion. Loading coils are toroidal inductors, which counter the effects of the phone pair's natural capacitance. While the coils are effective at flattening out the response within the voice band, the roll-off above 3.5 kHz is devastating, as shown in Figure 3.10-7.

Physically, load coil banks are long cylinders, with the individual donut-like coils stacked one on top of the other inside. They are typically placed at 3,000 (.9 km), 4,500 (1.4 km), or 6,000 (1.8 km) ft intervals along the phone cables. Generally, loading coils are found only on cables of greater than 3 miles (4.8 km) in length.

As we shall see, loading coils can create problems for the hybrids used in broadcast interfaces.

4-Wire Circuits

It is possible to purchase analog 4-wire circuits from telcos. These are used where it is desirable to maintain

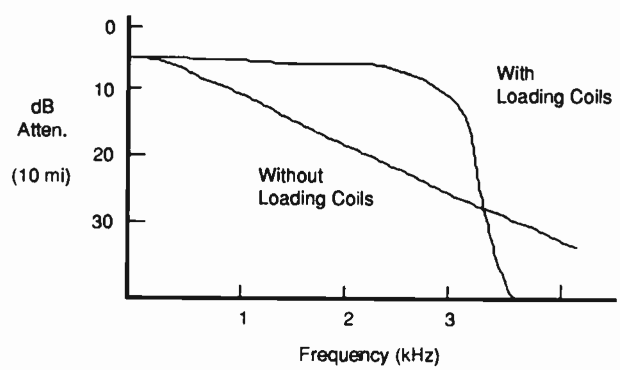


Figure 3.10-7. Frequency response with and without loading coils.

separation in the two speech paths. They are not dial-up, but rather end-to-end hardwired. This service has traditionally been used by television remote trucks for connection of remote production intercom systems. With the introduction of digital hybrid interfaces, use of this approach has been in decline. ISDN offers 4-wire capability at a lower cost and with fewer hassles, so it will probably supplant these analog lines over time.

Foreign Exchange (FX) Loops

FX provides local telephone service from a central office that is outside (foreign to) the subscriber's exchange area. If a station is located in the suburbs and the choke network central office is downtown, FX loops will be needed to connect your lines. When the phone is picked up, you get dial tone not from your local suburban CO, but from the downtown office. FX service is also sometimes used to extend your coverage into another city, so that people can call the station without paying a toll charge and calls can be made within that city without incurring toll charges. For instance, if the studio is in Cleveland and the goal is to serve listeners in Akron as if they were local, FX service could be the answer.

An FX loop is a 4-wire circuit with hybrids at each end, at each terminating central office. Since FX loops add an extra layer of hardware to the phone audio, they are another source of problems for on-air interfacing. They usually are engineered to have a few dB loss and they add to the impedance complexity of the line.

FX circuits are usually expensive and pose certain technical challenges. Since, as we will learn later in this chapter, hybrids are imperfect, a potential for a special kind of feedback called *singing* exists. This results from the inevitable leakage from the send to the receive ports at each hybrid. The phone people solve this problem by inserting a pad—anywhere from 5–8 dB is common.

Choke Networks

Most stations need special high volume exchanges for their contest and request lines. This requirement probably results from the days when aggressive program directors (PDs) desired the publicity that *burning out* a phone exchange would generate.

The choke network works by diverting calls beginning with the unique choke prefix around the local serving central office and sending them directly to the choke switching exchange, usually located downtown (see Figure 3.10-8). The phone company dedicates very few talk paths (wire trunks or special carrier equipment) to the task of connecting the caller's serving CO choke ports to the choke exchange. The usual switching and routing process is bypassed. Unfortunately, only a very limited number of paths are generally provided. In the densely populated Los Angeles area, for instance, only three connections exist from most central offices. In addition, the poorest facilities are often given over to the high volume service.

Generally, unless you are near the choke central office, the FX circuits previously described are employed to connect the choke CO to your serving CO. This is one of the reasons why choke circuits often have a lower level than standard lines. Because of their higher complexity, choke lines also usually have bumpier impedance curves, making good hybrid performance difficult to achieve due to the problem of finding appropriate balancing network values. This is especially a problem with simple analog hybrids.

In some areas, FX circuits are being replaced by internal call forwarding. This means that a published number is actually being software forwarded to a real number originating from your local serving CO. The main advantage to this approach is lower cost, since

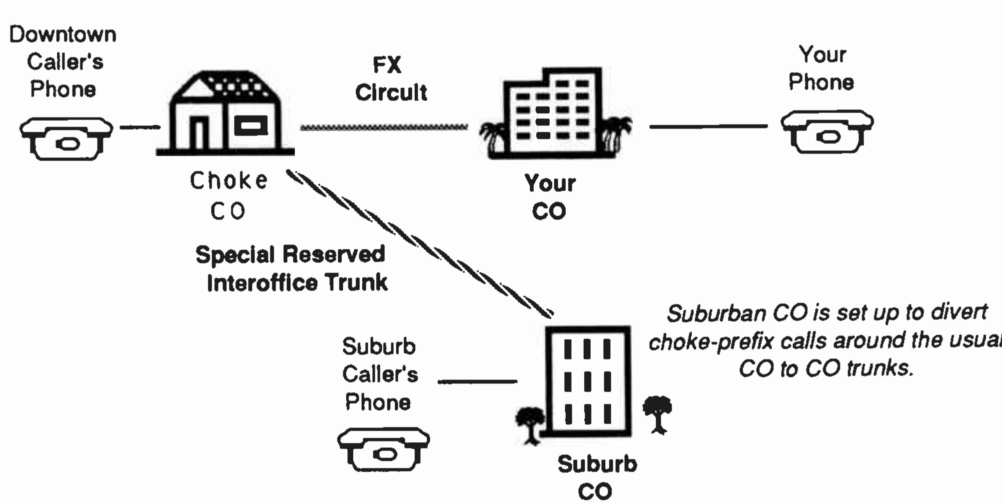


Figure 3.10-8. Typical choke network transmission path.

you do not have to pay the premium for the FX circuit. However, there usually is a smaller call-forwarding charge.

ISDN: Basic Rate Interface (BRI)

ISDN allows a direct digital connection to the telephone network. In addition to the quality advantages digital transmission offers for basic voice service, users may bypass the normal POTS speech coding methods and supply their own much better algorithms, such as those standardized by Moving Pictures Expert Group (MPEG). MPEG is an organization involved in standardizing audio coding. Another characteristic of ISDN important to broadcasters is that the B channels are true full-duplex, with absolutely no cross-connection between the send and receive signal paths.

ISDN is now widely available and is growing in popularity—mostly because of its value for high-speed Internet connectivity. Web surfers may implement direct digital links without the bottleneck caused by inefficient, slow modems. An ISDN BRI has 128 kbps raw capacity. Compare this to the speed possible with a 33.6 kbps modem and it becomes evident why the promise of ISDN creates so much excitement among people who need fast access to the net.

With a BRI line, you get two 64 kbps voice or data channels, called "B" or bearer channels, and one 16 kbps "D" or data channel on a single telephone pair (see Figure 3.10-9). The D data channel is the path between the central office and terminal equipment that is used for call set-up and status communication and is usually not available to the user.

The S and U Interfaces

The line from the central office is a single copper pair physically identical to a POTS line. When it arrives at the subscriber, this is called the "U" interface. The U interface converts to an S/T interface with a small box called an "NT-1." In the United States, NT-1 functionality is usually included in the terminal equipment. In Europe, the telephone company provides the NT-1. Only one NT-1 may be connected to a U interface, but as many as eight terminals may be paralleled onto an S bus.

Professional equipment should usually provide access to the S interface, making it possible to parallel multiple terminals. You can use either an external NT-1, or the equipment may have an internal NT-1 with both U and S/T connectors.

Terminal Adapters

A terminal adapter (TA) is the equipment that interfaces to the ISDN line, providing call set up and protocol conversion functions. A traditional TA has an ISDN connection on one end and one or two bit stream ports on the other, usually using the V.25 or X.21 connectors. Modern broadcast equipment combines this capability with the audio encoding equipment into one integrated unit.

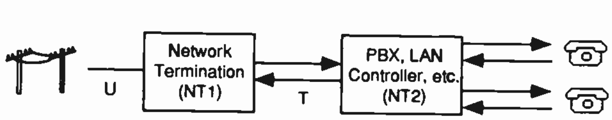


Figure 3.10-9. ISDN termination.

SPIDs

Service profile identification numbers (SPIDs) are only required when you are using the National I-1 ISDN protocol in the United States. This number is given to the user by the phone company and must be entered into the TA in order for the connection to function. SPIDs usually consist of the phone number plus a few prefix or suffix digits.

The intention of the SPID is to allow the telco equipment to automatically adapt to various user requirements by sensing different SPIDs from each type or configuration of user terminal. For instance, multi-button phones could retain function assignments when moving from line to line. In this case, the line number would probably not be used as the SPID. None of this matters with our application, but we must enter the SPIDs nevertheless. (Over time, it may be possible that a standard SPID could be used for all broadcast codec applications. A proposal that would allow this is being considered.)

If you are using the National I-1 protocol, your telco service representative must give you one or two SPID numbers for each line ordered. You will get one SPID for each B channel you need. Upon power-up, connection of the ISDN line or boot, the TA and the telco equipment go through an initialization/identification routine. The TA sends the SPID and, if it is correct, the network signals this fact. Thereafter, the SPID is not sent again to the switch. You must have this SPID number, and it must be 100% correct, or the system will not work. Do not let the installer depart without leaving your SPID number(s).

Directory Numbers (DNs)

Directory numbers (DNs) are the telephone numbers assigned to the ISDN line. You may be assigned one or two, depending upon the line configuration. If you have two active ISDN B channels, you will usually have two DNs. However, the physical channels are independent from the logical numbers. A call coming in on the second number will be assigned the first physical B channel, if it is not already occupied. Therefore, there must be some way for the TA to sort out which call goes to which channel/line. The DN is used for this function.

When a call rings in, it contains set-up information, which includes the DN that was dialed by the originating caller. The last seven digits are matched with the DNs programmed into the TA and the proper assignment is made. However, it is not usually necessary to explicitly enter them, as they are almost always contained within the SPID, and most TAs are smart enough to look there first. The only time a DN must be entered is in the very rare case where the last seven digits of the DN are not included somewhere within the SPID. When DNs are required, only the last seven digits need be entered.

Digital Long-Distance

Long-distance connectivity is routinely available in most parts of the United States from the big-three carriers: AT&T, Sprint and MCI. The "dial 1+" default carrier may be chosen at the time you order the line, just as with traditional voice lines. Also, just as with voice lines, you may usually choose a carrier on a per call basis by prefixing the number with the 1010XXX carrier selection code. You must dial the full number, including the 1 or 011 + country code following the prefix.

Here is a hot tip: You can save a lot of money by arranging a special plan with your long-distance (LD) carrier. When you use 1+ dialing without contacting your LD carrier, you are generally put into a standard rate plan that has the highest cost of any of the pricing tiers.

Some long-distance connections are limited to 56 kbps/channel. This arises from a quirk of the older telephone infrastructure. The channel banks that have been widely employed in the long-distance network have a native 64 kbps capability but rob the low order PCM bit on every sixth frame in order to convey supervision information (on-hook/off-hook and dial pulses). This limitation is becoming more rare as equipment is upgraded, but there is no way to know for sure in advance.

CSD and CSV

Recall that each ISDN BRI has two possible B channels. It is possible to order a line with one or both of the B channels enabled, and each may be enabled for voice and/or data use. Phone terminology for this class of service is circuit switched voice (CSV) and circuit switched data (CSD). (Both are in contrast to packet switched data (PSD) which is possible but irrelevant to this discussion.)

CSV is for standard voice phone service and allows ISDN to interwork with analog phone lines and phones. CSD is required for MPEG codec connections. Even though you may be sending voice, the codec bit stream output looks like computer data to the phone network.

Even for MPEG codec applications, you may want POTS speech capability, since some support this feature. Therefore, you may want to order CSV as well as CSD on one or both B channels. To get a line with one B channel to be used with either hi-fi or speech, you would request an ISDN BRI 1B+D line with CSV/CSD capability. For both B channels, you would order an ISDN BRI 2B+D line with CSV/CSD on both channels; if you do not need voice possibility on the channels, you want 2B+D with only CSD enabled.

Protocols

In a perfect world, all ISDN terminal equipment would work with all ISDN lines, without regard for such arcana as 5ESS, DMS100, CSV/CSD, SPIDs, etc. Unfortunately, the ISDN standard has been evolving for years and has only recently begun to settle down. And, sadly, there will remain different standards for the Unites States and Europe.

The telco network and the TA communicate via a *protocol*—the language the user equipment and the telephone network use to converse (on the D channel) for setting up calls and the like. This is where you will find differences, since the protocol depends upon the central office equipment and the standards that it follows.

In the United States, telephone companies use either AT&T 5ESS, Northern Telecom DMS100, or Siemens EWSD switches. Each of these can support the National I-1 protocol standard, which has been specified by Bellcore. However, both AT&T and Northern Telecom had versions of ISDN which pre-date the NI-1 standard and some switches have not been upgraded to the new format. There is also a newer NI-2 standard, but it is designed to be compatible with NI-1 for all of the basic functions.

In Europe, the common protocol is Euro-ISDN, following the ETS300 standards. It is an apparently successful attempt at having all of the European telephone networks use a single, compatible protocol. The telco authorities in most countries have adopted it already, with most of the rest planning to do so.

T-1 Digital Service

As with ISDN, T-1 is possible because an ordinary copper phone pair can carry a much wider signal than the 3.4 kHz required for a single voice conversation. Indeed, a pure metallic path of reasonable length is easily capable of passing frequencies in excess of 100 kHz. Thus, digitization and multiplexing can be used to carry a number of voice channels over a single pair of wires.

Introduction to T-1

To create the T-1 bit stream, 24 64 kbps DS-0 channels are assembled serially and the equivalent of another 8 kbps channel is added for synchronization (see Figure 3.10-10). Thus the ultimate data rate becomes 1.544 mbps, a rate also called DS-1. The signal is then converted into a digital bipolar bit stream in a special format called binary 8-zeroes suppression (B8ZS). The voltage is modulated between -3 V and $+3$ V.

Most LD carriers offer service on T-1 connected directly to their point of presence (POP). Because the LD carrier does not have to pay the usual fee to the local telco for routing over their CO and lines, the customer cost can be lower.

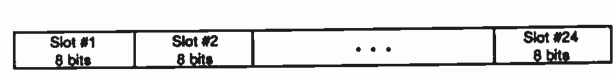


Figure 3.10-10. T-1 bit stream. 24 audio channels are transmitted sequentially.

Using T-1: The Customer Provided Equipment (CPE)

Despite the difference in capacity and service, T-1 arrives at the end user site as two conventional copper pairs: one for the data send and another for receive. The physical connector used to be a DB-15 type, but the current standard is the common RJ-48C, an 8-position modular plug. Figure 3.10-11 shows both types.

Here are the usual components of a terminal system for a T-1 circuit:

- *The CSU and DSU.* The T-1 line is first connected to a piece of equipment called the channel service unit (CSU). The CSU used to be considered part of the network, but is now almost always customer-provided and may also be merely included as an adjunct section in a complete T-1 interface solution. The CSU contains the last signal regenerator as well as a number of testing and maintenance features such as provision for loopback testing by the central office. It may also include a system to collect and report error statistics. The data service unit (DSU) handles the remaining digital housekeeping functions and data conversion from the bipolar T-1 format to standard serial data
- *The Multiplexer and Channel Cards.* The multiplexer, sometimes called a *channel bank,* is where the multiple voice (or data) channels are combined into the single bit stream required for T-1 transmission. Each voice channel is converted to and from digital using codecs. In order to simulate typical telco lines, talk battery is added, ringing voltage is generated and loop current is detected. Generally, multiplexers are constructed using a modular circuit card approach so that the available digital bandwidth may be configured as desired.

Many modern PBX systems and at least one broadcast on-air system are able to accept T-1 lines directly. This is a near ideal approach, since you get a low cost direct digital connection into the telco network. In this case, no multiplexer and channel cards are necessary, because the connection is made directly to the CSU/DSU. Some PBX equipment even incorporates the DSU.

T-1 and the Broadcast Interface

Generally, T-1 service appears to be a good idea for broadcasters, and many stations are using it successfully. However, be aware that some T-1 terminal equipment has problems in its analog conversion section, which cause the on-air hybrid interface to work very poorly with bad cancellation the result. Also keep in mind that, since all of your service will depend upon a single set of circuits, reliability could be reduced compared to individual analog lines. Consider having back-up circuits in place.

Primary Rate ISDN (PRI)

Primary rate ISDN has a data rate equivalent to T-1 circuits, providing 23B + D, or 23 64 kbps bearer channels and a 64 kbps D channel for control. (In Europe, PRIs have 31 bearer channels.) It is expected to replace T-1 eventually, since it speeds dialing and offers superior monitoring capabilities.

ADSL

Asymmetric digital subscriber lines (ADSL) promise connections at speeds of up to 3 mbps in the direction from the CO to the user. The upstream speed is limited to some much smaller value which is where the asymmetric part of the name comes from. An important advantage is the cost; it appears that this service may be priced at around the same level as ISDN BRI.

Initially, this technology was viewed by the telco industry as a way to compete with cable TV for the delivery of video services. Combined with an MPEG video/audio encoder, the bit rate offered by ADSL would permit full-quality National Television System Committee (NTSC) television. These projects now ap-

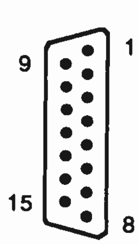
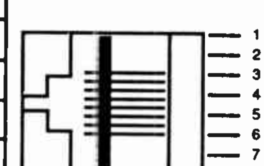
DB-15	PIN		PIN	RJ48-C
	11	R (Receive from Network)	1	
	3	T (Receive from Network)	2	
	9	R1 (Transmit to Network)	4	
	1	T1 (Transmit to Network)	5	
	N.A.	Optional Shield	7,8	
		(all others - no connection)		

Figure 3.10-11. T-1 connector pin-out. Either DB-15 or RJ48-C modular connectors may be used.

pear to be stalled and current efforts are being focused on high speed Internet connectivity. Since the Internet is a packet-based system with no bandwidth guarantee, the utility of this service for broadcast audio transmission is unclear.

Centrex

This service goes by various names, but the consistent principle is that the telco's CO equipment replaces customer-owned PBXs. Each phone set has a direct connection to the CO. The idea is to eliminate customer up-front costs and transfer maintenance responsibility to the telco. Varying requirements for numbers of lines or phones can be accommodated without customer equipment upgrades. Centrex is declining in popularity but seems to remain popular with universities.

Features in Centrex rely upon flashing the switchhook and the use of the normal dialpad keys, generally an awkward and confusing situation for users. This problem may be solved with ISDN Centrex, as this permits very sophisticated phones to be used with all of the usual PBX features.

Cellular Telephone

Cellular extends the dial-up network to many places where a wire connection would not be considered practical. Cellular transceivers operate in the 800 MHz range and automatically select the appropriate frequency from among the 666 FM channels assigned for this service. Low power is used so that the frequencies can be re-used in adjacent areas. The mobile phone varies its power according to the level of signal received at the base location. A useful feature for on-air use of a cellular phone is the signal strength meter provided on some units. Some phones also allow you to see the send power value. Often, the antenna's pattern is quite directional due to its position on the vehicle, so moving around while observing the level indication can help make remotes sound better. For fixed remotes, a Yagi antenna can be used with its benefits of higher gain and directionality. At 800 MHz, Yagis are very compact.

Most equipment designed for use with wired phone lines can be connected to cellular phones using an adapter provided by the phone manufacturer. Intended for laptop computer modems and portable fax machines, these adapters provide an interface to any broadcast equipment that can connect to a phone line. Units especially designed for broadcast use have provisions for audio input and output for direct connection to microphone mixers and the like.

Some new digital cellular systems have the capability to transfer data via a special interface. Unfortunately, the bit rate is limited to only 14.4 kbps—not sufficiently fast for digitized audio. The impetus from the Internet may cause cellular vendors to offer higher bit rate phones in the future, permitting broadcasters to use them for high-fidelity remotes.

A downside of the new digital phones is that speech quality may be poor. This results from the very low bit rate used by these systems and the extreme compression methods that are required to shoehorn audio into the channel.

FCC Regulations

FCC requirements for connecting equipment to phone lines are outlined in Title 47 of the Code of Federal Regulations (CFR), Part 68: Connection of Terminal Equipment to the Telephone Network. The CFR can be ordered from the Government Printing Office.

PBX AND KEY SYSTEMS

Now that we know a bit about the nature of the phone network, we can explore what happens after the lines become ours. We will want to use some of what the phone people refer to as CPE. That is all of the equipment connected to the phone line after the official demarcation point. We will survey the various styles of PBX systems available both for general office and on-air use, followed by a look at systems designed specifically for studio application.

Private branch exchanges (PBXs) are found where there is a need for a large number of extension phones. PBXs are miniature central office exchanges, allowing local phones to call each other as well as access trunk lines for incoming and outgoing calls. PBX systems often have a number of specialized features for call routing and control. Traditionally, PBX systems have used only single-line phone sets as terminals, with special functions like transferring and conferencing accessible by flashing the switch hook or by using the tone pad in a special way. Most PBXs now have available feature phones, which can button-access individual lines as well as provide numerous other advanced functions. Sometimes these systems are called key systems after the old multi-key 1A2 phones. (Why phone engineers called buttons keys remains a mystery.)

Modern Telephone Systems

While the systems are tremendously varied, most have in common that the cable from each phone set to the common equipment conveys:

- Power to operate the phone
- A two-way data path to signal user actions from the set to the switch and operational and display status from the switch to the set
- The speech audio.

Here are the usual approaches phone manufacturers employ for wiring and communication:

- *All Digital.* The most advanced systems use a pure digital bit stream for both voice and data. The phone set contains the codec for conversion to-and-from the analog and digital domains. The pure digital approach is used in the AT&T System 85, in the

Northern Telecom Meridian family, in the newer Mitel systems with the *Superset DN* phones and in the digital version of the NEC NEAX, among many others. The Siemens *Office Point* system claims to use standard ISDN protocol between the sets and the common equipment

- *Separate Pair per Function.* The early electronic phones used a separate pair for each of the three functions, and thus required three (or more) pairs. The AT&T Merlin system used this design. The center pair is the audio; another pair is for the serially transmitted control and display data and another handles the phone's power requirements
- *Two-Pair, Phantom Power.* This used to be the most common approach, but is now fading, as pure digital designs have become cost-effective. The AT&T Spirit system the popular NEC and TIE systems and many others use this approach. Talk and data each use one of the two pairs. The power is applied between the two pairs similar to the method used for phantom powering condenser microphones in recording studios. A transformer at each end of the audio pair permits the phantom power to be added. The data pair will probably use resistors to obtain a center tap, rather than transformers since the data signal has a dc component which could not pass through a transformer.
- *Two-Pair, Power not Phantom.* Some two-pair systems put the data on one pair and the audio on the other. Power may be on the data pair or on the audio pair. In the latter case, the audio pair resembles a central office line so that the phone ports may be universal: either single-line sets or feature phones can be plugged-in without hardware changes in the PBX. At least one of the Panasonic systems uses this technique. The center pair, again, is generally the audio
- *Data Over Voice.* The analog Mitel *Superset* phones use a unique scheme that requires only one pair for all three functions. The data is amplitude shift modulated onto a 32 kHz carrier *over voice* and then the combined voice and data are ac coupled across the dc power voltage.

Interfacing to PBX Phones

It is usually possible to interface to PBX phones for on-air use. However, this is best reserved for casual phone use such as for the occasional request or contest winner call. For applications where phone calls are a significant programming element, it is usually better to consider the specialized on-air systems from the broadcast-oriented manufacturers.

One reason is that the hybrid interface cannot determine when a new call is selected, so it can not adjust its null to the new line before the conversation starts. (However, since the hybrid can null on voice during conversation, null will be achieved in perhaps four seconds. This is acceptable if only a portion of the call is to be aired, as is common with on-air requests, contest winner calls and the like.) Another shortcoming of the direct-to-electronic phone approach is that the line switching clunk is not muted, although this is not a problem when calls are not aired directly and sequentially.

Another potential problem is audio quality. The primary impediment is usually noise, most often the result of the data signals cross-talking into the audio. Buzz from the power supply sometimes finds its way into the audio. Often, frequency response is limited by too small line coupling transformers or from other causes. Poorly designed digital systems may suffer from quantization and aliasing noise and distortion. Few PBX manufacturers publish specs on audio performance. Since, clearly, this is of importance to those of us who need to get decent quality from phones for on-air use, we'll want to make sure that the audio is at least reasonable. When choosing a new PBX, ask the phone system dealer for audio performance data or arrange to conduct at least a few simple tests yourself.

Direct Connection to the Skinny Wire

When the phone system uses the separate-pair approach previously described, the center two wires on the modular plug are usually the audio path. Since the phone's control functions stay active even when these connections are broken, it is possible to intercept the audio signal here for feed to the interface. Most broadcast interfaces provide a loop-through connection, which feeds the phone line back out when it's not active. Thus, the unit may be series connected with the audio pair. That way, you have normal telephone function preserved when the interface is not in use. When the interface is active, the phone serves merely as a controller, with no audio reaching the phone's network or handset. Wiring the hybrid's on/off functions to the console's switching logic accomplishes automatic operation.

When the phone uses the two-pair phantom approach previously described, the audio is again likely to be present on the inner pair and may be intercepted for interfacing use if the dc connection is maintained. One way to do this is to provide a bypass for dc with inductors. Two µH has proven acceptable in experiments performed on some phone systems As shown in Figure 3.10-12.

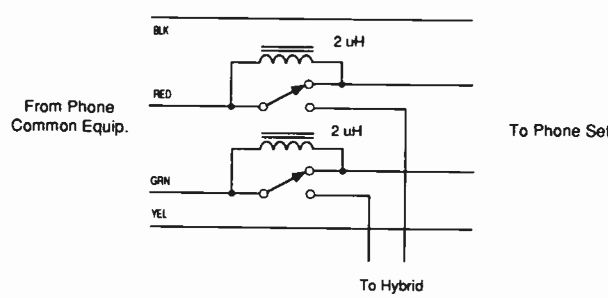


Figure 3.10-12. If the center two wires on many electronic phone systems convey audio, they may be used to feed broadcast equipment. The inductors bypass power to the phone set when the studio interface is active.

Special System Ports: Faux CO Lines

Since fax machines and modems need connections that look like central office lines, many systems provide ports for this use. They may be connected to broadcast interfacing gear as if they were CO lines. Sophisticated PBXs have programming features that allow these ports to be configured in various and potentially useful ways. For example, they may be set up for private line ringing (when a given incoming CO line rings, the call may be directly sent to the selected port). Unfortunately, with most PBX systems, awkward operation may result, since the only way to move a call from a phone set to the port may be to transfer it using multiple button punches, rather than the usual simple place-on-hold-and-pick-up-elsewhere operation. Taking calls in sequence on-air may be extremely difficult or impossible. Figure 3.1013 illustrates one possible solution.

Speakerphone Tap-Off

One way to get low cost interfacing is to take advantage of the switching-type interface that many phone set internal speakerphones provide. The procedure is to tap off the speaker with a transformer and pad to the console's required input level. You may continue to use the phone's internal microphone or you can provide an external send audio source to substitute for the phone's internal microphone. Again, you will certainly need a pad and probably a transformer. The input feed must be set so that appropriate switching action and proper send levels are obtained.

Handset Adapters

Adapters are available that plug into the phone set's handset modular jack and convert the microphone and earpiece signals into a signal that emulates a standard CO line. While useful in some applications, this approach is likely to offer a lower quality feed because the phone set's network remains in the signal path causing impedance bumps and other problems.

Intercepting the Serial Data Stream

Why can't we just emulate an electronic phone set by generating and decoding the phone system's serial data? It does seem that this would be a good solution. However, phone system manufacturers insist on keeping their data protocols a deep secret. That means that broadcast manufacturers are unable to design direct emulation equipment. Of course, even if we had the protocols, there is the problem of accommodating the dozens of communication methods employed by PBX designers.

1A2 Key Systems

While nearly all stations have gone to high-tech PBXs for the business office, many on-air installations continue to rely upon 1A2 key systems. Key systems offer the advantage of providing a direct metallic connection to the CO line. That means that no frequency response error, noise, distortion or time delay is introduced. Often, these issues are not fully considered in the design of the more complex business phone systems. In addition, costs are favorable, and full schematics and other documentation are readily available.

Leading from the key service unit (KSU) to each phone is a thick cable with 50 conductors (25 pairs). The *tip/ring* pair carries the telephone audio. As mentioned, these are direct connections to the telco CO lines. The *A leads* tell the key system which lines are in use and also signal a hold condition. Selecting a line causes a connection to be made in the phone set from the A lead to another wire, the *A-common*. The A lead is normally at -24 Vdc and A-common is at ground potential, so when a line is selected, the A lead goes from -24 Vdc to ground. If the A lead is broken before the tip/ring is disconnected, the system puts the

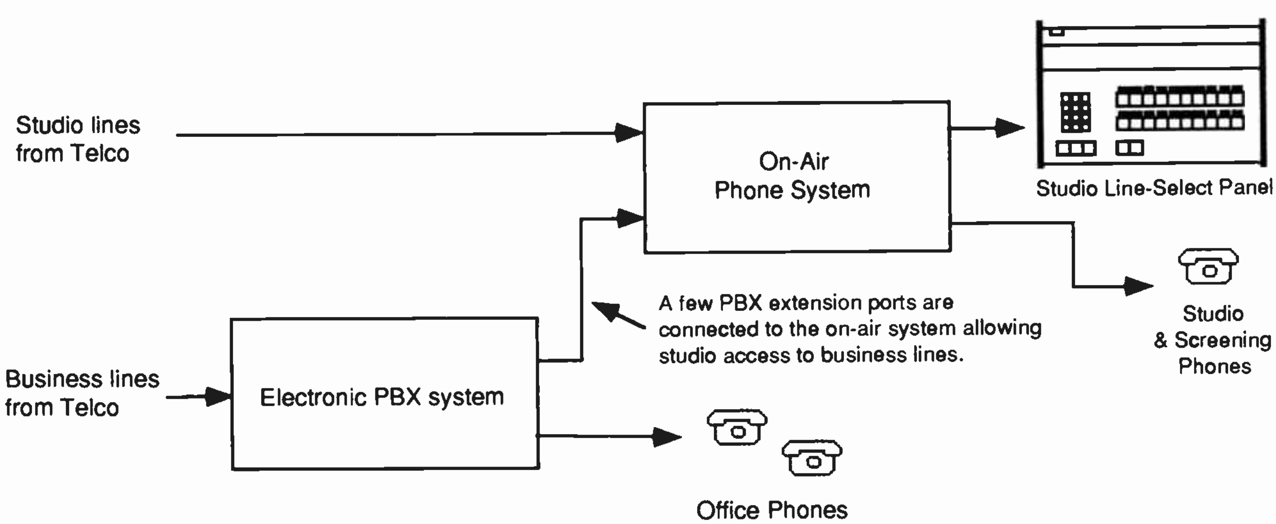


Figure 3.10-13. One way to integrate the on-air system with the station business phone system. Ports intended for single-line phone sets are used as input to the on-air system.

Table 3.10-2

Telephone color code and 1A2 key system assignments. The pin numbers indicated are for the Amphenol "Blue Ribbon" connectors used to terminate 25-pair cables.

Pin #	Wire Color	9 Line 1A2	5 Line 1A2
26	WHITE/BLUE	Line 1 tip	Line 1 tip
1	BLUE/WHITE	Line 1 ring	Line 1 ring
27	WHITE/ORANGE	Line 1 A	Line 1 A
2	ORANGE/WHITE	A circuit common(gnd)	A circuit common(gnd)
28	WHITE/GREEN	Line 1 lamp ground	Line 1 lamp ground
3	GREEN/WHITE	Line 1 lamp	Line 1 lamp
29	WHITE/BROWN	Line 2 tip	Line 2 tip
4	BROWN/WHITE	Line 2 ring	Line 2 ring
30	WHITE/SLATE	Line 2 A	Line 2 A
5	SLATE/WHITE	Line 9 A	A circuit common(gnd)
31	RED/BLUE	Line 2 lamp ground	Line 2 lamp ground
6	BLUE/RED	Line 2 lamp	Line 2 lamp
32	RED/ORANGE	Line 3 tip	Line 3 tip
7	ORANGE/RED	Line 3 ring	Line 3 ring
33	RED/GREEN	Line 3 A	Line 3 A
8	GREEN/RED	Line 8 A	A circuit common(gnd)
34	RED/BROWN	Line 3 lamp ground	Line 3 lamp ground
9	BROWN/RED	Line 3 lamp	Line 3 lamp
35	RED/SLATE	Line 4 tip	Line 4 tip
10	SLATE/RED	Line 4 ring	Line 4 ring
36	BLACK/BLUE	Line 4 A	Line 4 A
11	BLUE/BLACK	Line 7 A	A circuit common(gnd)
37	BLACK/ORANGE	Line 4 lamp ground	Line 4 lamp ground
12	ORANGE/BLACK	Line 4 lamp	Line 4 lamp
38	BLACK/GREEN	Line 5 tip	Line 5 tip
13	GREEN/BLACK	Line 5 ring	Line 5 ring
39	BLACK/BROWN	Line 5 A	Line 5 A
14	BROWN/BLACK	Line 6 A	A circuit common(gnd)
40	BLACK/SLATE	Line 5 lamp ground	Line 5 lamp ground
15	SLATE/BLACK	Line 5 lamp	Line 5 lamp
41	YELLOW/BLUE	Line 6 tip	
16	BLUE/YELLOW	Line 6 ring	
42	YELLOW/ORANGE	BL, AG, or spare	BL, AG, or spare
17	ORANGE/YELLOW	SG, LK, or spare	SG, LK, or spare
43	YELLOW/GREEN	Line 6 lamp ground	
18	GREEN/YELLOW	Line 6 lamp	
44	YELLOW/BROWN	Line 7 tip	
19	BROWN/YELLOW	Line 7 ring	
45	YELLOW/SLATE	B or B1	B or B1
20	SLATE/YELLOW	R or R1	R or R1
46	VIOLET/BLUE	Line 7 lamp ground	
21	BLUE/VIOLET	Line 7 lamp	
47	VIOLET/ORANGE	Line 8 tip	
22	ORANGE/VIOLET	Line 8 ring	
48	VIOLET/GREEN	Line 9 lamp ground	
23	GREEN/VIOLET	Line 9 lamp	
49	VIOLET/BROWN	Line 8 lamp ground	
24	BROWN/VIOLET	Line 8 lamp	
50	VIOLET/SLATE	Line 9 tip	
25	SLATE/VIOLET	Line 9 ring	

line on hold. The *lamp-leads* light the phone's line buttons with 10 Vac from the KSU's power supply and are returned via the *lamp grounds.* The standard color codes and pinout are given in Table 3.10-2.

The Evolving Phone

As time goes on, probably all but the most inexpensive systems will use the purely digital approach. As we've seen, these systems are difficult to interface to, but perhaps over time protocols will become standardized and maybe even based on ISDN. If this happens, broadcast interface manufacturers may be able to provide equipment that could directly connect to the PBX in place of, or in series with, the studio phone set.

Computer Telephony Integration (CTI)

With such a system, the PBX manufacturer provides complete documentation on an interface that can provide control of all of the important aspects of phone switching, including call set-up and routing functions. A standard data port is provided so that outside vendors may supply systems to work in concert with the phone equipment. These open PBXs may eventually offer a universal method for broadcast equipment to coordinate with the station's office phone system.

Another approach is to build a PBX using special cards and software installed in a standard PC. Systems of this type would use the Windows NT operating system along with other standard PC software components such as a database server to provide a very sophisticated package of features. It is possible, for instance, to dial using database name look-up on a networked PC. Ironically, most CTI systems use analog phone sets.

BROADCAST INTERFACING

This section describes the techniques necessary to achieve the best possible result from the phone-to-broadcast shotgun marriage.

One-way Interfacing

There is often a need to take audio from a phone or broadcast in only one direction at a time (newsroom phoners are a common application). If there is no requirement for a two-way conversation, a simple interface using a QKT will do. Formerly available from the phone company, this small box was permanently wired into a phone instrument or line and provided a quarter-inch (12.7 mm) phone jack output for feeding a line-level signal to a console or recorder input.

Since the QKT is nothing more than a transformer, a capacitor and a zener diode limiter, you can make your own (see Figure 3.10-14). The capacitor provides dc blocking so that the transformer does not become saturated with the phone line's dc potential. In order for the coupler to hold the line by drawing loop current, eliminate the capacitor and use a transformer that can withstand the loop current without producing distortion. (One such a transformer is the SPT117 from Prem Magnetics.) When sending audio into the phone line, remember audio level should be limited to −9 dBm. The QKT had back-to-back zeners for this purpose; you may want to add them to your homemade interface if you expect audio levels to get out of hand. Of course, commercial units are available that are a little fancier than the simple device described here. Some offer auto-answer and disconnect capability.

When using a coupler, it is most convenient to have the telephone instrument on-line and equipped with a push-to-talk switch on its receiver. This is because the phone's receiver has to be off-hook while a feed is coming in; the switch turns off the receiver's mouthpiece microphone when it is not depressed, thus insuring that noise from the studio side will not be included in the recording. Since this coupler works in both directions, it can be used to send audio down the phone as well—useful in the production studio for letting

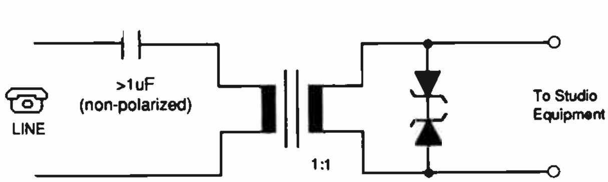


Figure 3.10-14. Simple one-way-at-a-time interface. The capacitor is for dc isolation and is not required when a transformer which can sink loop current is used. The zeners are chosen to properly limit transmission levels to the required −9 dBm.

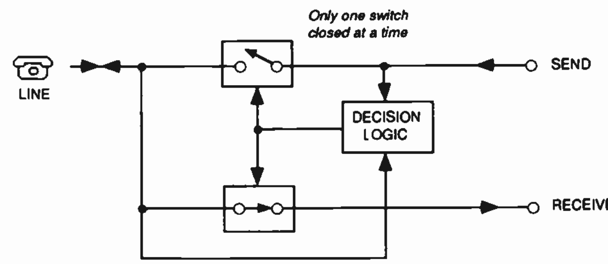


Figure 3.10-15. Switching interface allows two-way conversation, but only one way at a time.

clients hear their commercial masterpieces before they go into the control room.

When hooking up to a multi-line phone, connect to a point where the tip/ring is present after line selection. The most convenient place is usually right at the phone network. Use headphones to find the spot.

Two-way Interfacing

The simple coupler's limitations become apparent when it is necessary for the caller to hear the announcer and the audience to hear the caller simultaneously. A more sophisticated method is needed because of the requirement to have isolated send and receive audio signals.

Switching

This is what you get when you connect a speakerphone to your console input. No commercial broadcast interface uses this technique, which uses gain switching to keep the send audio from appearing at the receive output. Two electronic switches or voltage controlled amplifiers are used in such a way as to ensure that either the send or the receive path is closed at any given time, but never both simultaneously (see Figure 3.10-15). A decision circuit compares the send and receive levels, with the direction of transmission being determined by the relative signal strengths.

The disadvantage of the switching technique is its uni-directional nature. The caller cannot be heard while the announcer is speaking, and noises in the studio can sometimes cause a caller to disappear momentarily, especially on weak calls.

The Hybrid

Hybrids were invented long ago to separate the send and receive signals from the common two-way phone pair. Early hybrids were made from transformers with multiple windings. Nowadays, most hybrids are made with active components and are known as *active hybrids.* Both circuit types use the same principle and achieve the same effect.

In Figure 3.10-16, the first op-amp is simply a buffer. The second is used as a differential amplifier; the two inputs are added out-of-phase (subtracted). If the phone lines and the *balancing network* have identical characteristics, then the send signals at the second differential amp will be identical, and no send audio will appear at the output.

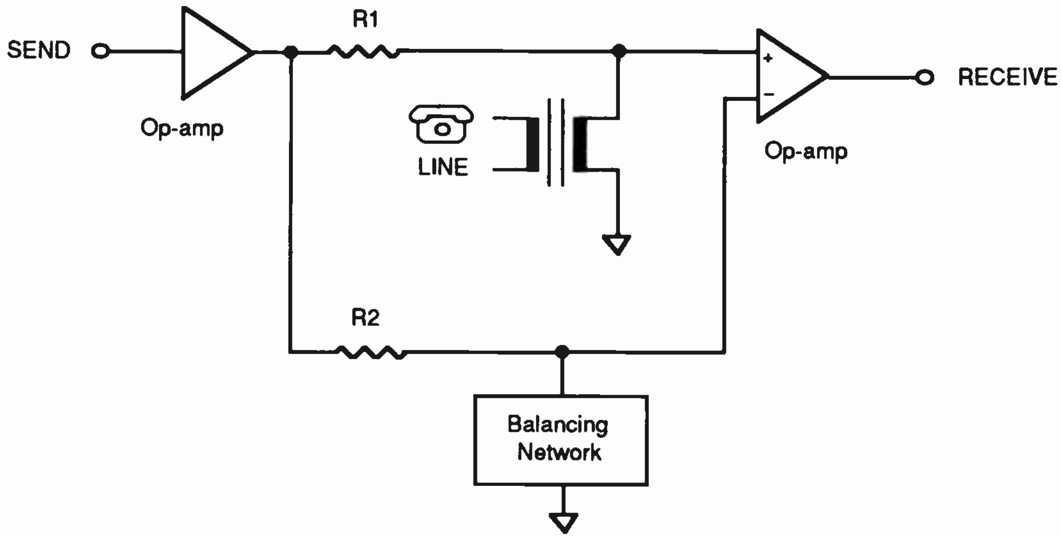


Figure 3.10-16. Op-amp hybrid. The second op-amp is used as a differential amplifier to perform the required subtraction for nulling.

The balancing network is a circuit consisting of capacitance, resistance and sometimes inductance, forming an impedance network. Depending on the hybrid's application, this circuit can be very simple or it can be comprised of a large number of components and have a very complex impedance characteristic.

R1 and the phone line form a voltage divider, as does R2 and the balancing network. If the phone line and balancing network are pure resistances, then, clearly, the phone line and the balancing network must have the same value in order for the signals at the differential amplifier to have the same amplitude and for complete cancellation to occur.

The phone line, however, is not purely resistive, but rather is complex impedance, causing both the amplitude and phase to vary as the send signal frequency varies. Two-to-four wire converters, transformers, repeaters, T-carrier systems and other telco systems are responsible for significant impedance bumps. Loading coils also usually have a deleterious effect on the performance of hybrid interfaces since the coils can create resonant peaks and phase anomalies in the phone line's impedance curve which are difficult to null out.

Only when the impedance of the balancing network is the same as the phone line, and the signals at the differential amplifier are matched in both amplitude and phase, will full cancellation of the send signal be achieved. Otherwise, leakage results—the scourge of hybrids.

Because the phone company's requirements are not generally too stringent, they usually use a simple network with compromise values of resistance and capacitance. Their goal is to get an average of about 12 dB rejection, with 6 dB acceptable on difficult lines—just enough to prevent feedback in a system with back-to-back hybrids. When the situation calls for better performance, modules with a number of R and C elements that can be switched in or out are employed, the switches being set to match the network to a particular line.

Broadcast Hybrid Application

In broadcast application, the studio mixing console combines the output of the hybrid and the announcer's microphone audio, as illustrated in Figure 3.10-17. As discussed previously, the hybrid output consists of both the desired caller audio and the undesired leakage—(the announcer audio but *phase-shifted because of the phone line's reactance*). If the amount of leakage is too great and the phase shift too extreme, the announcer sound will suffer degradation as the original and leakage audio combine in and out of phase at the various affected frequencies. When this occurs, the announcer sounds either hollow or tinny as the phase cancellation affects some frequencies more than others. Another effect of too little transhybrid loss is that feedback can result from the acoustic coupling created when callers must be heard on an open loudspeaker. Yet another problem can occur when lines are to be conferenced; when the gain around the loop of the multiple hybrids is greater than unity, feedback singing will be audible. So a hybrid will be useful for broadcast only when leakage is kept acceptably and consistently low.

The plots of phone line impedance vs. frequency and phase shift shown in Figure 3.10-18 are the result of measurements performed on phone lines at a radio station in the Midwest. They indicate the wide variation seen on typical telco lines as provided to broadcasters. The lines with smooth curves have impedance characteristics that could be emulated with a simple resistor-capacitor (RC) combination. These lines would work fairly well with a simple hybrid, since an RC balance network would match the impedance characteristic closely enough to make the cancellation of send audio at the hybrid output good enough to prevent coloration of the announcer audio.

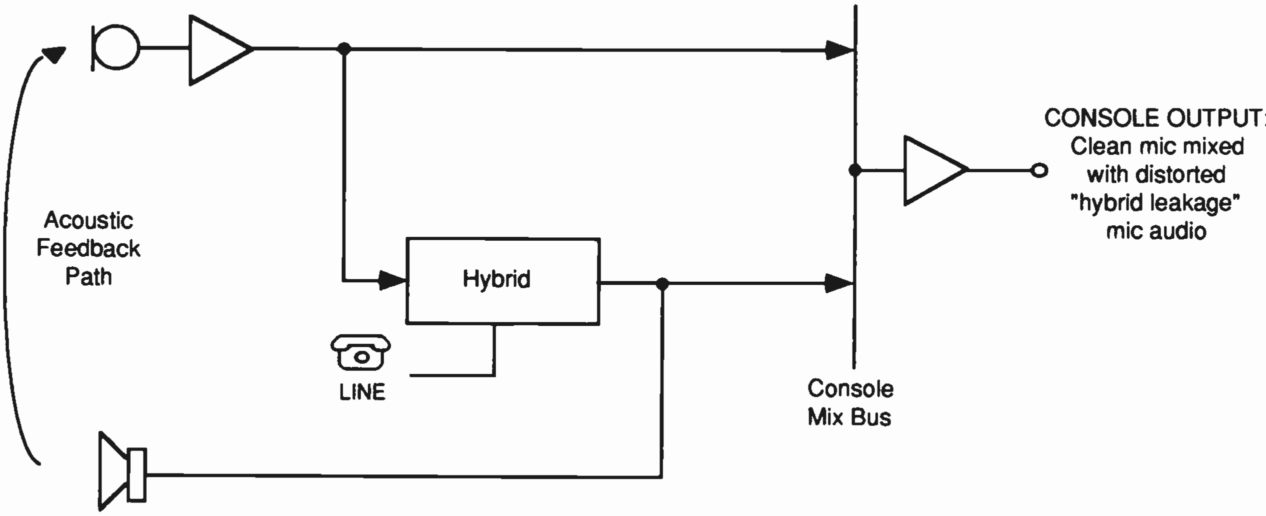


Figure 3.10-17. Block diagram of typical studio arrangement with telephone hybrid. Announcer audio is combined with hybrid output, potentially causing problems with announces voice distortion. The acoustic path is a possible source of audible feedback.

Those other lines are quite another story! While it would theoretically be possible to construct a balance network to match the difficult lines, practical considerations usually keep this approach from being used. The impedance characteristic required is too difficult to produce using resistors and capacitors. If the hybrid is to be switched among a number of lines, the line characteristic would have to be consistent from call-to-call and nearly the same impedance curve.

Digital Signal Processing Hybrids

Digital signal processing (DSP) offers a very powerful and effective technology to improve hybrids. DSP is the process of operating on analog signals that have been converted into the digital domain. Since the signals are numbers, mathematical operations can be performed to manipulate them before being returned to analog. Complex processing functions either impractical or impossible to be done with analog circuit elements are achievable in DSP.

With the DSP hybrid, natural simultaneous conversation is possible without distortion of the announcer audio. To accomplish this, the announcer and caller audio signals are digitized and processed in a system that makes use of a specialized DSP microprocessor. The digital hybrid incorporates software programmed to perform the hybrid cancellation function. The technique, *convolutional least mean square adaptive filtering*, is capable of very accurate synthesis of the required balancing transfer function for maximum nulling (see Figure 3.10-19). Unlike resistor/capacitor analog schemes, the adaptive filter can create the com-

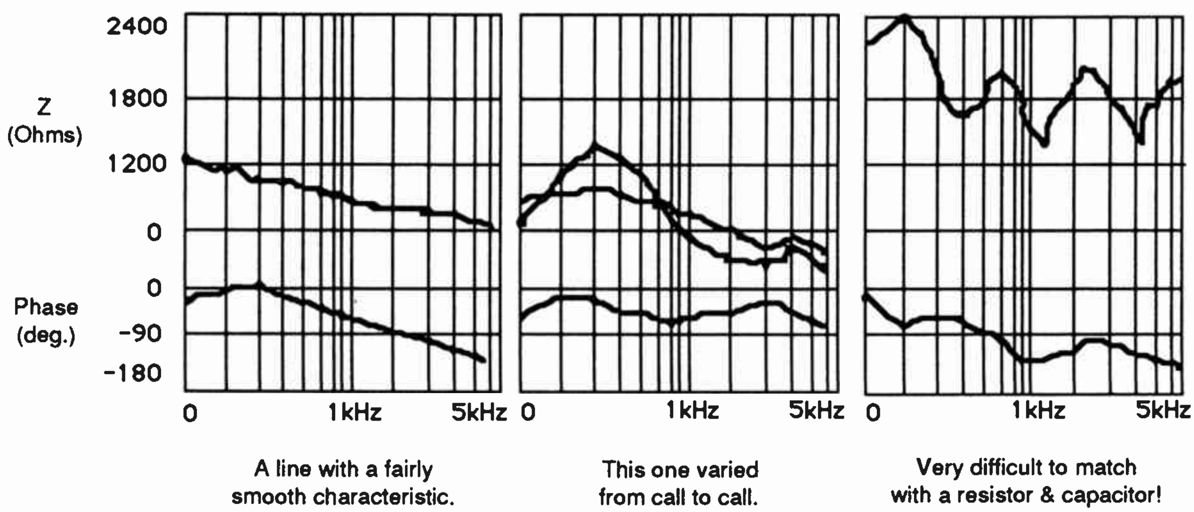


Figure 3.10-18. Impedance vs. frequency curves for some typical phone lines.

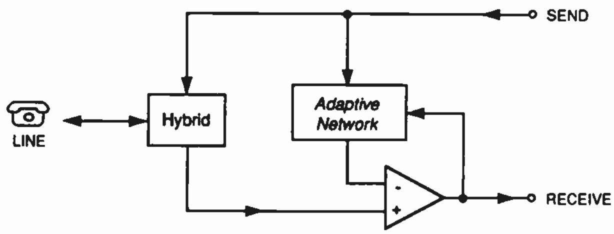


Figure 3.10-19. In the DSP hybrid, the digital balancing network continuously adjusts to the phone line impedance characteristic. When the adaptive network transfer function is identical to that of the phone line, perfect cancellation is achieved. Since the adaptive network is a digital filter than can create almost any required curve, performance is superior to the analog hybrid alone.

plex multiple break-point impedance vs. frequency curves required by difficult-to-match phone lines. The send and receive signals are constantly compared in a feedback loop with the leakage becoming an error control signal which drives adjustment of the *digital balancing network.*

The performance advantage of the digital hybrid technology is striking. On a typical phone line with a fairly smooth impedance curve, an analog hybrid might attain 15–20 dB transhybrid loss. A digital hybrid will likely produce 40 dB or better transhybrid loss. On lines with difficult impedance curves, the analog hybrid's performance will usually be so poor as to prevent its use, while a digital hybrid would perform acceptably.

When a call is initially established, a brief mute/adaption period provides an opportunity for the system to adjust to the phone line prior to the call going on air. The caller hears a noisy tone, but none of this tone is heard on the air since the output is muted. This has the incidental benefit of removing the line switching clunk. Adaption continues as the conversation proceeds, using voice as the reference signal.

While in the digital domain, other operations in addition to the hybrid adaptive balancing can be performed. Automatic gain control (AGC) can take advantage of digital techniques to significantly improve upon the functions implemented in analog. For instance, cross coupling to the hybrid section is possible in order to avoid the output AGC, confusing hybrid leakage with low level caller audio and inappropriately increasing gain. AGC may be smartened in other ways, as well. An adaptive floating expansion threshold, for example, improves noise-gating quality.

Evaluating Hybrid Performance

The amount of hybrid rejection—the transhybrid loss—directly affects the on-air audio and is the most critical measure of hybrid quality. The true test of hybrid performance is determined by measuring the amount of rejection across the entire audio frequency range, preferably with pink noise as a test signal at the send input. Any hybrid with an adjustable R and C balance network can produce high rejection at a single frequency, since both phase and amplitude at a single frequency can be adjusted for good cancellation.

Another thing to keep in mind—although the two are related, the transhybrid loss is not the same as the observed difference between the caller level and the leakage at the hybrid's output. That is because the typical phone call is maybe −20 to −25 dBm (on choke lines, even lower) and the send level (to the caller) from the hybrid should be −10 dBm. That means that the hybrid has to use up 10–15 dB of its transhybrid loss just to get even. The remainder becomes the observed difference.

Other important performance characteristics include S/N ratio, distortion and (for a digital unit) number of bits in the audio path. The operation of the dynamic functions—the AGC, noise gate and override ducking—make a significant contribution to a hybrid's effective performance.

Combining the Hybrid and Switching Techniques

This is the method used in nearly all commercial interfaces. The hybrid produces as much send-to-receive isolation as can be achieved. Then a ducking or *override* function causes the *dynamic* rejection to be greater than the hybrid alone can produce. When send audio is present, the receive gain is reduced. Thus, leakage also is minimized. However, since the level from the phone is also reduced when the announcer is speaking, there is a sacrifice of full-duplex operation. A user adjustment in the control signal path permits variation of the amount of receive ducking, allowing full duplex operation when the hybrid alone produces sufficient rejection, or speakerphone-like operation whereby the caller is turned almost completely off when the announcer speaks. As a practical matter, this control is usually set to provide the minimum amount of ducking which provides adequate send-to-receive leakage suppression.

ISDN For Studio Call-In Talk Systems

ISDN can provide a direct digital connection to the POTS analog network, so it can be used to enhance the quality of on-air calls. A call set-up message is sent from the customer equipment to the network to tell it to switch into POTS interworking mode. (This is in contrast to when an ISDN line is used with MPEG codecs. In that case, the line may be carrying voice signals but in a format that is incompatible with POTS phones. Instead, the network is providing a transparent end-to-end digital path.)

The cost of ISDN service is not a barrier. With ISDN lines costing about the same as analog in most parts of the United States. An ISDN BRI, with two channels, costs about twice as much as a POTS line. (Pricing varies depending on the telco but ranges from a 20% discount to a 30% premium. The average is probably around a 10% premium.

Broadcast interfaces may use either BRI or PRI. A simple interface for the newsroom could use a single BRI. Even sophisticated multiline systems could use

BRIs, with enough of them to achieve the desired number of lines. While PRIs would seem to be a more technically appropriate solution for a multiline system, BRIs may be more cost effective, more readily available and able to provide a measure of redundancy. A system using PRI or T-1s may be able to share lines among a number of studios, with connections to both hybrids and codecs.

ISDN Lines Are Inherently 4-Wire

As we have learned, analog lines use a single pair of wires for both signal directions, mixing the send and receive audio. This causes the famous leakage problem—where the announcer's audio is present on the interface output, instead of the desired caller only audio. Digital circuits inherently offer independent and separated signal paths.

While DSP based hybrids applied to the problem of separating the send/receive signals are dramatic improvement over analog systems, ISDN enables further improved performance. This is because it offers a fully independent path for each speech direction. In the case where both ends of a connection are digital, there is no mixing whatsoever. In the call-in application, the far-end from the studio will still be 2-wire, so the audio paths will not be fully independent and a digital hybrid function will still be necessary to cancel residual leakage. But moving the studio side connection away from mixed analog can help tremendously because it provides the hybrid a much better starting point.

Better Digital-Analog Conversion Quality

The codecs used in telephone central offices are not as good as the converters commonly used in audio equipment. Fidelity is not an important consideration when designers choose parts for this function. In a professional interface for studio application, we are able to design with much better converters than available in the telco's equipment. Noise-shaping functions permit a larger word-length converter to provide significantly better distortion and S/N performance.

In all digital installations, the phone interface can maintain a digital path all the way. Audio Engineering Society/European Broadcasting Union (AES/EBU) can be provided on the interface to accomplish the connection to the studio gear.

Lower Noise

As digital circuits, ISDN lines are not susceptible to induced noise. Analog lines are exposed to a wide variety of noise and impulse trouble-causers as they move across town on poles and through your building. Hum is the main one, given the line's proximity to transformers and ac power lines, but there are also sources of impulse noise from motors, switches and other sources. Digital lines convey the bits precisely and accurately from the network to your studio equipment without any perturbation—so the audio remains clean.

Call Setup and Supervision are Better

Analog lines use a strange mix of signaling to convey call status. Loop current drop and returned dialtone signal that a far-end caller has disconnected; blasts of 100 volts at 20 Hz mean someone wants you to answer. Why should we be using a mechanism designed to bang a gong against a metal bell to transmit network status information in the 1990s? ISDN uses a modern digital approach to controlling calls and conveying status information about them. The sophisticated transactions on the D channel are able to keep both ends of a call accurately informed about what is happening.

ISDN call set-up times are often a few tenths of milliseconds, enhancing production of a fast-paced show. Perhaps more importantly, when a caller disconnects while waiting on hold, the ISDN channel communicates this status change instantly. This contrasts with the usual 11 second delay on most analog lines. One of the most common complaints of talk hosts is that when they go to a line where they expect a caller to be waiting, they are met instead with a blaring, annoying dialtone. The chance of this happening with an ISDN line is reduced to near zero.

Another common error is when a talent goes to punch up a line that looks free, but is actually just about to begin ringing and connects to a surprised caller. This condition results from the delay in the ring signaling, which comes from the nature of the analog line's ringing cadence. This is much less likely with ISDN because the ambiguous status period is eliminated.

Levels

ISDN does not have the FCC-mandated -9 dBm send level limit. Audio may be adjusted to fill up the digital word, resulting in higher send signal volume.

Reduced Feedback During Multi-line Conferencing

When conferencing is required on 2-wire circuits, very good hybrids are needed to separate the two audio paths to add gain in each direction. When the gain around the loop exceeds unity, there is the possibility of feedback singing. Since the conference path usually includes four AGC functions, the hybrid must be sufficiently good to cover the additional gain that may be dynamically inserted. Because of the 4-wire nature of ISDN, the hybrid function is more effective and more reliably so across a variety of calls. That means more gain can be inserted between calls before feedback becomes a problem.

Line Monitoring

Since there is a full-time connection between the central office and the terminal on the D channel, it is possible to detect when a line is a not working. On an analog line, one discovers a problem only from a failed attempt to use the line.

The ISDN Broadcast Interface

Most of the functions performed by an ISDN interface are similar to that of an analog DSP hybrid, but there are some differences, both in the required functions and in the implementation of the common features.

Send/Receive Separation. This is the traditional hybrid function provided by broadcast telephone interfaces. Despite the fact that ISDN lines naturally have two independent send and receive paths, it is still necessary to provide additional functions to further reduce leakage. The reason is that almost all calls will originate with telephone sets connected via 2-wire analog lines, and so there will still be a mixing of both speech directions.

Acoustic Coupling Reduction. There is often an acoustic path between the received caller audio and the send audio signal. This results from having a loudspeaker in the studio that produces sound that couples into the microphones. When the talent use headphones for monitoring callers, this is not a problem. But sometimes it is not practical to convince guests to wear headphones, and television stations generally do not want talk show talent to wear earplugs. In these cases a combination of adaptive cancellation and dynamic gain reduction will reduce the coupling electronically.

High-grade Digital-to-Analog Conversion. When an analog connection to studio equipment is required, pro-grade converters can be used to provide much better quality than the usual telco conversion. At minimum, 16 bits should be used, but 18–20 bits may not be excessive.

Sampling-rate Conversion. When the studio connection is via a digital AES/EBU channel, no analog-digital conversion is required, but it will be necessary to adapt the sampling rate of the telephone network to the studio rate. telco sampling rate is 8 kHz, and studio equipment will usually operate at 32, 44.1 or 48 kHz. A process is required to perform the required up-and-down sampling, while suppressing aliasing and reconstruction audio components.

Automatic Gain Control. As with POTS hybrids, this function should be provided on both the send and receive audio paths. On the send side, it is necessary to smooth the wide level variations that arise from usual studio practices. Talent are used to having on-air processing take care of level variations and are generally not very careful at riding gain. On the receive side, AGC is essential to deal with the very different levels that can result from the many types of phone sets and telco analog network components.

Dynamic Equalization. With phone sets having a very wide variety of microphone characteristics, a multiband automatic equalizer helps callers have a reasonable spectral consistency.

Caller "Ducking." As with POTS hybrids, this can serve to reduce residual leakage. However, since ISDN hybrids have much better inherent transhybrid loss, this feature will be used mostly to satisfy a programming aesthetic requirement, reducing the level of the phone audio when the host talks and allowing her an automatic control over a caller who wants to carry on. This is a matter of taste; some talent and programmers prefer no ducking so that hosts and callers can conduct heated exchanges without impediment, while others want to exercise control.

Caller ID. ISDN naturally conveys caller ID information. This is transmitted instantly in the setup message and is much faster than the 1200-baud modem method used in analog caller ID.

Conference Linking. With two B channels available on one BRI line, broadcast interfaces will be dual units, making possible high quality conferencing between the two potential callers. Some systems will probably support larger numbers of conferenced callers.

BROADCAST ON-AIR SYSTEMS

With phones an important part of programming at many stations, systems to enable convenient, high quality on-air integration of phone conversations are essential.

On air phone systems are specifically designed for use in the broadcast studio environment. While many business phone systems offer similar functions—line selection and status indication, conferencing—they are generally awkward to operate in an on-air environment and may have other limitations such as the audio quality flaws described earlier.

While the phone network would not be considered to be a high fidelity source, it clearly does not help to degrade it further by adding additional noise, distortion or frequency response impediments. For that reason, broadcast phone systems are designed with these issues and other specialized requirements in mind. For example, a broadcast phone system output should be free of inappropriate switching sounds, and air talent should be able to access and manipulate lines live without any pops or clunks being audible to listeners.

Ergonomic Requirements

Line selection and other functions must be performed intuitively and with a minimum of hassle. Unlike a telephone set, broadcast line selection panels have large illuminated buttons. To avoid operator confusion, features are limited to those necessary for on-air application. One such example is panels that drop into an open position in the studio mixing console so that the line selection buttons are located near the channel on/off, fader, and audio switching functions.

Conferencing Capability

Most broadcast systems allow any number of lines to be switched to air, even if only a single hybrid is present. But, unless you are blessed with excellent phone lines, you will want additional hybrids with each connected to the other through a multiple mix-minus arrangement. That way, it will be possible to have amplification between callers. Without multiple hybrids, callers might have difficulty hearing each other, since you are at the mercy of the telco-delivered line level.

Special Features

Desirable features for an on-air phone switching system include:

Busy/unbusy. To prepare for a contest, all lines may be busied-out and then returned to readiness after the contest has been announced.

Automatic next line selection. Pressing the next button picks up the line that has been holding the longest. If no line has been holding, the longest ringing-in line is selected.

Call length timer. Displays call duration time.

Held caller timer. This tells which line has been holding longest and for how long.

Integration of On-Air Systems with PBXs

To interconnect the on-air system with the front office PBX, there are a number of possibilities.

Segregate the studio and office phone lines. Ports from the PBX configured to look like CO lines feed an input or two on the studio system so that that calls taken by the receptionist can be put on the air.

Route all lines through the PBX. The studio lines are programmed in the PBX to be forwarded to the ports that feed the studio system. Some audio degradation may result.

Simply parallel the two systems. With no cross-coupling of line status information, there could be trouble if a line is inadvertently picked up on one system while the other is being used.

Route the on-air lines through the broadcast system. Possible if the broadcast system brings out a loop-through connection. This scheme prevents PBX phones from picking up active on-air lines.

Improving Phone Audio Quality

Whether extracted from analog or digital lines, due to its limited frequency response and fairly high distortion, the audio from the phone has the poorest quality of our on-air sources. Thus, it generally pays to make telephone audio less of an earsore so that it does not stand out more than is necessary from other program material.

If the phone network is a digital system, why do phones still sometimes sound pretty awful on the air? The main problem is that phone engineers never designed the systems with a connection to full fidelity broadcast systems in mind. The 8/13 bit quantization scheme used for phone speech coding results in less than high fidelity. Often, the problem lies in the specific implementation rather than in any inherent shortcoming in the standard or the technology. One important quality limitation results from the anti-aliasing and reconstruction filters in the codecs. These filters usually have an ultimate roll-off of around 35 dB. Audio above the 4 kHz Nyquist frequency will alias and appear in the 300 Hz–3.4 kHz band as distortion. Thus, typical codecs have distortion of 2–3% from aliasing. The strange raspy noise that seems correlated with the speech sometimes heard on a telephone circuit is a result of the effects of this kind of distortion combined with audible quantization errors.

Also the codec filters generally use switched-capacitor technology, which tends to be fairly noisy. Some newer codecs avoid the switched-capacitor problems by employing the same *delta-sigma* over-sampling and digital decimation concept used for high performance digital audio conversion, but these are only rarely found in telco central office equipment.

What can we do? An ISDN connection solves half of the problem, since at least one of the telco's codecs is bypassed. We still have the other end to contend with, and the majority of broadcast connections will remain analog. Fortunately, there are some remediation possibilities. Filtering, equalization, gating and dynamics compression are the primary tools. Most of the commercial hybrid interfaces have at least some of these processes built-in.

Filtering

On a dial-up phone line, there is very little audio above 3.4 kHz—but there is noise. Thus, a filter with a very steep roll-off above the telephone passband will reduce phone line noise significantly without affecting conversation audio. The low-end can be improved as well. Low-frequency hum is often a problem—usually 60 Hz mixed with its second harmonic, 120 Hz. Thus, it is often a good idea to have a sharp roll-off starting at 200 Hz or so.

Equalization

An equalizer used to shape the frequency response of the phone line within its audio bandwidth can result in marked improvements in perceived quality. A typical phone line has an excess of energy at around 400 Hz and considerable roll-off at both the top and bottom ends of its passband, so the idea is to compensate by adding gain at both. Boosts at 2.5 kHz and 250 Hz and a cut at 400—500 Hz with a parametric equalizer will help achieve better sound. Since every phone line is different, the ear is usually the best instrument to evaluate the results.

When it is not possible or practical to make custom adjustments, an adaptive multiband EQ can be an effective tool. The principle is much the same as implemented in broadcast transmission processors. Audio is filtered into multiple bands, and an automatic gain adjustment is performed on each spectral segment. Given the limited frequency range of telephone calls, three bands are sufficient.

Noise Gating

Another effective processing device is the expander or noise gate. These devices may be used to reduce gain between the words of a conversation, thus making phone line noise less objectionable. On extremely noisy lines, however, the gating action can make noise more distracting by causing it to come and go with the words. In such cases, it might sound better to leave the gate off and let the noise remain present at a constant level. A unit with variable threshold and duck ratio can be adjusted so that the optimum compromise may be achieved between the benefit of reduced noise and audibility of the effect.

Dynamics Compression

Levels on phone calls vary widely, and it is not uncommon to see levels range from −40 to near −4 dBm as calls are switched into a given line. A compressor helps to smooth the levels. An AGC that maintains a constant compression ratio regardless of average gain reduction produces more consistency. Freeze gating is also important, so that gain does not increase during caller speech pauses.

Mix-Minus: Getting the Send Audio Feed

The feed-to-caller signal has come to be referred to as *mix-minus*, so called because it is often the *mix* of all of the console's active inputs *minus* the phone hybrid's output (see Figure 3.10-20). A mix-minus feed is necessary because the hybrid will create a feedback path if it is forced to chase its tail. Usually, the mix-minus is a mix of only the studio microphones, but it may sometimes include other audio that is to be sent to the phone such as contest sound effects from cart machines.

To create the required signal in simple installations, a feed taken from the main announce microphone may be all that is necessary. The patch send output available on many consoles is precisely what you need. In installations where multiple microphones are to be used, a combiner of some sort is required. This may be a small outboard mixer or a homemade op-amp summer or even a resistive combiner. Better consoles offer special purpose busses that may be used for mix-minus, often with provision for selective switching of sources into the phone feed. If you need to modify an older console that does not have special buses, a device (made by Henry Engineering) accomplishes the mix-minus by subtracting the hybrid audio from the console program output with a differential amp scheme. This unit generates a mix-minus signal true to its name—all sources except the phone itself will feed the phone.

Recording Phone Calls

Some stations may want to record calls for later playback. One technique is to have the mix-minus go to one track of a stereo tape machine, while the other channel gets the hybrid output with the caller audio. The result is a two-track tape with the announcer and caller separated. To play back, the console's input mode is set to mono; the relative balance, if need be, can be adjusted upon playback. The production department can use its tape to facilitate extraction of contest squeals.

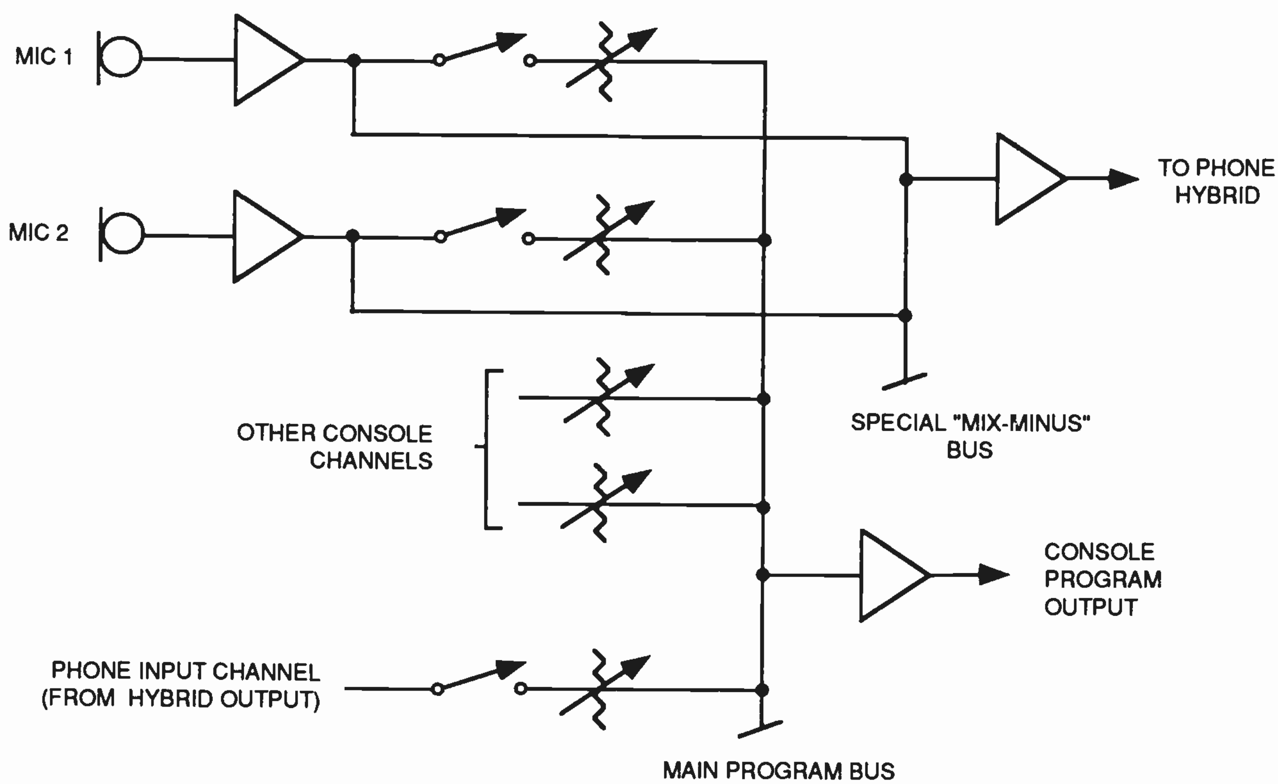


Figure 3.10-20. Simplified studio block diagram shows the mix-minus required for hybrid feed.

Talk-Show Screening Software

In its simplest form this personal computer software lets a talk show screener/producer communicate to the air talent who's on the line waiting to talk. It replaces the paper pieces on the window system employed for years at many talk stations. The better packages offer a number of convenient features: display of liner messages and other information, storage of caller data for demographic analysis and remote operation via modem.

An Ethernet or serial port on the broadcast system can let the computer display reflect current line status. New software enables laptop computers to extend full control capability and status display to a remote site, and modern systems even permit this function to be conveyed over the Internet.

ISDN: HI-FI REMOTES ON DIAL-UP LINES

ISDN makes high quality remotes possible with dial-up convenience. Convenient, reasonably priced studio-quality audio from almost anywhere in the world is now possible. The enabling technologies are digital telephone services like ISDN and audio compression or coding algorithms. Products offering this capability have burst onto the market in the last few years, and broadcasters have enthusiastically embraced the programming possibilities created by the new capability.

BROADCAST CODECS

Broadcast codecs have evolved rapidly over the past few years. Most are now single-box solutions that include an ISDN TA interface and a number of selectable coding algorithms. Some portable units even include a mixer for multiple audio inputs and outputs. Most are full duplex, with provision for transmitting and receiving simultaneously, and most offer the ADPCM G.722 and perceptual MPEG Layers 2 and 3 coding algorithms. Some offer a feature to allow dialing to POTS phones for low-grade voice communications. State-of-the-art systems include an auto-dial feature that adjusts the codec section settings, such as bit rate and transmit and receive coding choices, as well as the numbers for the codec you wish to dial.

In MPEG modes, many codecs permit bidirectional serial data at 9.6 kbps to be transmitted simultaneously with the audio. End-to-end parallel contact-closures offered by many codecs may be used to control recorders and other devices. Since codecs are inherently digital devices, it is only natural that AES/EBU digital inputs and outputs are usually available. Sample rate conversion is generally available on both input and output paths.

For many remotes, a receive-side mixer is required to combine the mix-minus signal from the studio with the local audio (see *Dealing with Delay*, below). In some cases, you will want to have two outputs: one for the talent, which can include cueing audio, and another for the public, who listens to PA loudspeakers. This is a common feature in portable units that are designed to be used in the field. Another valuable feature for remote applications is an input audio limiter to prevent digital nasties when the program signal peaks instantaneously, as might be the case on remotes hosted by excitable sports announcers.

The J.52 Protocol

While codec manufacturers have been remarkably successful at making their products inter-operate, it is often necessary to manually adjust a unit at one end or the other to a compatible mode. The J.52 standard addresses this problem by including information in the transmitted bit stream which identifies the details of the encoding method being used, allowing the receiver to automatically conform. J.52 also standardizes channel bonding.

ADPCM Coding

Adaptive delta pulse code modulation (ADPCM) pre-dates MPEG perceptual coding. It has been around as an international standard the longest and is probably the most widely used system. ADPCM is much simpler than the perceptual methods but suffers from poorer audio performance. It has the benefit of low cost and the unique advantage of low delay.

The most popular method, G.722, was invented in the late 1970s and adopted as a standard in 1984 by the Consultative Committee for International Telephony and Telegraphy (CCITT), a division of the United Nations. The technique used is Sub-Band ADPCM, which achieves data reduction by transmitting only the difference between successive samples. G.722 does this in two audio frequency sub-bands: 50 Hz–4 kHz and 4 kHz–7 kHz.

G.722 has a frequency response extending to 7 kHz at 56 or 64 kbps. Unless there is no alternative, it should be used only for voice feeds, as music transmitted via G.722 has a distinct fuzzy quality. It is good also for cueing and intercom channels. Only two bits are allocated per sample for audio frequencies above 4 kHz—sufficient for conveying the sibilance in voice signals but not very good for intricate musical sounds. Also, the *predictor model* used to determine the step size in the adaptive function is designed only for speech.

G.722 has the lowest delay of all popular coding methods, about 20 msec. For this reason, it often used as a return channel so that the round-trip delay is reduced, even when a higher fidelity method is used for the on-air feed.

Statistical Recovery Timing

G.722 uses a procedure called *statistical recovery timing* (SRT) or *statistical framing* to lock the decoder to the data stream. This procedure is specified in ANSI standard T1.306-1989.) The process usually happens instantaneously but can take up to 30 seconds.

The locking can be sensitive to audio present on the G.722 path, as it relies upon the properties of the audio bit stream itself. Some audio material and tones can prevent lock from ever happening. Silence is the most reliable signal for locking, and undistorted voice is usually acceptable. The most common problems are with sine tones and distorted voice or music signals, in which case, removing or lowering (to around −12 dB) the audio for a few seconds will generally cause lock to occur. In very rare cases, it may be necessary to disconnect and redial. Other strange effects may be observed. Tones and noises may be present before locking occurs, and some continuous audio tones may cause momentary unlocking.

Perceptual Coding

The broadcast world has been transformed by the introduction of perceptual audio coding techniques. Applying perceptual coding methods, it is possible to pass studio-quality 15 or 20 kHz bandwidth audio over ISDN channels.

Demystifying MPEG

By far, the most popular perceptual coders rely upon techniques developed under the MPEG umbrella. About a decade ago, when the CD had just been introduced, the first proposals for audio coding were greeted with suspicion and disbelief. There was widespread agreement that it would not be possible to satisfy golden ear listeners while deleting 80% or more of the digital audio data. In response, the MPEG was formed, and since 1988 the group has been working on the standardization of high quality low bit rate audio coding. Two standards have been completed: MPEG-1 (coding of mono and stereo signals at sampling rates of 32, 44.1 and 48 kHz) and MPEG-2 (ISO/MPEG IS-11172: coding of 5 + 1 multi-channel sound signals and low bit rate coding of mono and stereo audio at sampling rates of 16, 22.05 and 24 kHz). Today almost all agree not only that audio bit rate reduction is effective and useful, but that the MPEG process has been successful at picking the best technology and encouraging compatibility across a wide variety of equipment.

The MPEG process is open and competitive. A committee of industry representatives and researchers meet to determine goals for target bit rate, quality levels, application areas etc. Interested organizations that have something to contribute are invited to submit their best work. A careful, double blind listening test series is then conducted to determine which of the entrant's technologies delivers the highest performance. The subjective listening evaluations are done at various volunteer organizations around the world that have access to both experienced and inexperienced test subjects. Broadcasters are the most common participants with recent test series conducted at the BBC, the CBC, NHK. Finally, results are tabulated, a report is drafted and a standard is issued.

In 1992, this process resulted in the selection of three related audio coding methods, each targeted to different bit rates and applications. These are the famous layers. In 1997, another algorithm, advanced audio coding (AAC) was added to the MPEG standard.

Acoustic Masking

All of the MPEG codecs rely upon the celebrated *acoustic masking principle*—an amazing property of the human aural perception system. When a tone—called a *masker*—is presented at a particular frequency, we are unable to perceive audio at nearby frequencies that are sufficiently low in volume. As a result, it is not necessary to use precious bits to encode these inaudible, masked frequencies. In perceptual coders, a filter bank divides the audio into multiple bands. When audio in a particular band falls below the masking threshold, few or no bits are devoted to encoding that signal, resulting in a conservation of bits that can then be used for the bands where they are needed.

MPEG Layer 2

MPEG Layer 2 is the world's most popular perceptual coding method. It is the preferred choice for applications where greater than 120 kbps/channel is available, such as satellite links and high capacity terrestrial paths such as Primary ISDN or T1 channels. Layer 2 is the method used for satellite television audio and for many other applications such as hard disk storage. It is also used for European Eureka 147 terrestrial digital broadcasting.

Layer 2 offers a *joint stereo* technique to improve coding efficiency with stereo signals. The Layer 2 joint stereo mode uses an intensity coding method. This process has high coding power and is quite effective; however, it may impair stereo separation on some program material as audio above about 3 kHz is combined to mono and panned to one of seven positions across the stereo stage, at lower bit rates.

MPEG Layer 3

MPEG Layer 3 is perfectly matched to the bit rates available on ISDN BRI lines, permitting full FM broadcast quality. Full fidelity 15 kHz mono is possible on a single ISDN B channel and very near CD-quality 20 kHz stereo is achievable using both ISDN B channels. Until equipment supporting the new MPEG AAC standard arrives, Layer 3 is the most powerful method available to broadcasters. It is widely supported in broadcast codec equipment from a number of manufacturers.

MPEG Layer 3 uses a number of advanced techniques to achieve its power:

Psychoacoustic Masking. The audio is divided into 576 frequency bands. First, a polyphase filter bank performs a division into the 32 main bands, which correspond in frequency to those used by the less complex Layer 2. Filters are then used to further subdivide each of the main bands into 18 more. The resulting bandwidth of each sub-band is 27.78 Hz. A 32 kHz sampling rate allows very accurate calculation of the masking threshold values. Sufficient frequency resolution is available to exceed the width of the ear's critical bands (100 Hz below 500 Hz; 20% of the center frequency at higher frequencies) across the audible spectrum, resulting in better hiding of noise than would otherwise be possible.

Redundancy Reduction. A Huffman coding process accomplishes redundancy reduction. Values that appear more frequently are coded with shorter words, whereas values that appear only rarely are coded with longer words. This results in an overall decrease in the data rate with no degradation, since it is a lossless reduction scheme.

Bit Reservoir Buffering. Often, there are some critical parts in a piece of music that cannot be encoded at a given data rate without audible noise. These sequences require a higher data rate to avoid artifacts. Layer 3 uses a short time bit reservoir buffer to address that need.

Ancillary Data. The bit reservoir buffer offers an effective solution for the inclusion of such ancillary data as text or control signaling. The data is held in a separate buffer and gated onto the output bit stream using the bits allocated for the reservoir buffer when they are not required for audio.

Joint Stereo. A joint stereo mode different from that in Layer 2 permits advantage to be taken from the redundancy in stereo program material. The encoder switches from discrete L/R to a matrixed L + R/L − R mode dynamically, depending upon the program material.

MPEG AAC

The MPEG-2 AAC system is the newest audio coding method selected by MPEG and become an international standard in April 1997. It is a fully state-of-the-art audio compression tool kit that provides performance superior to any known approach at bit rates greater than 64 kbps and excellent performance relative to the alternatives at bit rates reaching as low as 16 kbps.

The development of AAC began when researchers became convinced that significant improvements would be possible by abandoning backward compatibility to the earlier MPEG layers. The idea was to start fresh and take the best work from the world's leading audio coding laboratories. Fraunhofer Institute, Dolby, Sony and AT&T were the primary collaborators. The hoped for result was *International Telecommunications Union (ITU)-R indistinguishable quality* at 64 kbps per mono channel. This was a fairly daunting requirement because it requires that no test item fall below the perceptible, but not annoying threshold in controlled listening tests. The test items include the most difficult-to-encode audio known to researchers—isolated pitch pipe, harpsichord and glockenspiel recordings, among others. The thinking was that if a coding system passes this requirement, it will almost certainly perform well with normal program material. Pop or western classical music is tremendously easier to encode.

Compared to the previous layers, AAC takes advantage of such new tools as temporal noise shaping, backward adaptive linear prediction and enhanced joint stereo coding techniques. AAC supports a wide range of sampling rates (8–96 kHz), bit rates (16–576 kbps) and from one to 48 audio channels.

The AAC system uses a modular approach. An implementer may pick and choose among the component tools to produce a system with appropriate performance-to-complexity ratios. Three default profiles have been defined, using different combinations of the available tools:

Main Profile. Uses all tools except the gain control module. Provides the highest quality for applications where the amount of random accessory memory (RAM) needed is not constrained.

Low-complexity Profile. Deletes the prediction tool and reduces the temporal noise-shaping tool in complexity.

Sample-rate Scaleable (SRS) Profile. Adds the gain control tool to the low complexity profile. Allows the least complex decoder.

AAC is the first codec system to fulfill the ITU-R/EBU requirements for indistinguishable quality at 128 kbps/stereo. It has approximately 100% more coding power than Layer 2 and 30% more power than the former MPEG performance leader, Layer 3.

Choosing the Coding Method Most Appropriate to Your Application

The following chart compares some of the important characteristics of G.722, Layer 2 and Layer 3.

One thing that should be apparent from Table 3.10-3

Table 3.10-3
Audio Coding Comparisons

	Layer 3	Layer 2	G.722
Method	Perceptual + Huffman	Perceptual	ADPCM
Audio Freq. Response/mono	15/20 kHz*	8/10 kHz**	7 kHz
Audio Freq. Response/stereo	15/20 kHz*	20 kHz	7 kHz
Delay at 32 kHz/mono	280 msec	—	20 ms
Delay at 48 kHz/mono	240 msec	150 msec	20 ms
Delay at 32 kHz/stereo	450 ms	—	20 ms
Delay at 48 kHz/stereo	340 ms	220 ms	20 ms
20 ms	MS Matrix	"Intensity Coding"	—
ISO Target Bit Rate	64 kbps/channel	128 kbps/channel	N/A
Coding "Power"	12:1	6-8:1***	4:1
Bands	576	32	2
Frequency Resolution (48 kHz)	42 Hz	750 Hz	—

* 15 kHz at 32 kHz sample rate; 20 kHz at 48 kHz sample rate.
** 8 kHz at a 56 kbps network rate; 10 kHz at 64 kbps.
*** 12:1 in intensity joint stereo mode.

is the trade-off between delay and audio performance. Layer 3's excellent audio performance requires a significant delay, because some of its power comes from the ability to analyze the audio over a relatively long period. Layer 2 requires the next longest delay, and G.722 has minimal delay.

The most flexible broadcast codecs permit the coding mode for the send and receive paths to be independently chosen, so the choice may be optimized for the specific requirement of each direction.

Dealing with Delay

All perceptual coders have too much delay for talent on remote to hear themselves via a round-trip loop. Therefore, a special mix-minus arrangement is required—exactly the same as has been used with satellite linked remotes for years. The principle is this: The remote talent does not hear himself via the studio cue return. Rather, his microphone is mixed locally with a studio feed that has everything but the remote audio thus the mix-minus designation. The announcer gets in his headphones a non-delayed version of himself and a slightly delayed version of all of the studio pieces.

To save money and hassle, callers are usually received at the studio, rather than at the remote site. In this situation, phones need to be fed to the remote talent so that they can hear and respond to callers, and the phone callers can hear them. In many cases, the remotes are sufficiently distant that the station cannot be monitored for the caller feed. Even if it could, the profanity delay would be a problem, since the talent needs to hear the phone pre-delay. Instead, the talent hears callers via the return path. As before, this return is fed with mix-minus: a mix of everything on the program bus minus the remote audio.

As for the second half of the equation, the callers hear the talent because the remote feed is added to the telephone mix-minus bus. This should be no problem if you have a setup that permits selective assignment to the phone mix-minus. The most common problem with this arrangement is a result of a phone hybrid with too much leakage combined with the system delay. If the hybrid isn't doing a good job of preventing the send audio from leaking to its output, the special remote send mix-minus is corrupted. Remember, if any of the announcer audio from the remote site is returned via the monitor feed, it will be delayed by the digital link, causing an echo effect. The answer is to make sure you have the best possible hybrid with the maximum transhybrid loss. If it has variable override (caller ducking), you could increase the amount when these remotes are in progress.

The round-trip delay in a typical remote broadcast may be reduced by using the G.722 algorithm for the return cueing path and MPEG for only the on-air direction.

Cascading Codecs

Coder cascading is an active field of investigation among algorithm designers, standards organizations and users.

Some of what we do know:

- Some International Consultative Committee for Radio (CCIR) tests have demonstrated that one pass of Layer 3 at 56/64 kbps can be cascaded with two-five passes of Layer 2 operating at high (112 kbps+/ per channel mono; 192 kbps+ joint stereo) bit rates with good results
- Informal tests at Telos with two passes of Layer 3 transmitted via Zephyr codecs have proven successful, with listeners noticing no audible degradation—even on difficult CDs
- One user has reported that two passes of Layer 3, followed by one pass of sedat, is acceptable (L3 in joint stereo mode).

SEDAT

The goal is to get as much coding headroom as possible at each stage. This is achieved by:

- Using the most possible bits at each stage—the least crunching—and/or
- Using the more powerful coding method of those available at each stage.

Here is some practical advice:

- Use coders only where necessary. Consider the alternatives at each stage. With the cost of hard disk capacity falling, is it really necessary to crunch at this point?
- Use the maximum bit rate you can afford at each stage. Hard disk recorders and other studio systems often have an option to adjust this. For very critical work, remember that some codecs may be used in a mode where a mono program is split over two digital network channels
- Use Layer 3 or AAC on low bit rate channels.

The staff at Fraunhofer Institute who developed the Layer 3 algorithm have introduced a computer based perceptual coding analyzer. This device has the potential of making objective measurements a reality and may help us learn about the effects of cascading with various coding methods and bit rates.

Mixed MPEG Layer 2 and Layer 3 Signal Chains

What about the case where you will be using L2 and L3 together in a signal chain? It turns out that the two methods are nicely complementary.

At low bit rates, Layer 3 gets more signal-to-mask margin than Layer 2. This is why it performs better in the low bit rate tests. It accomplishes this by using a filter bank with more bands, 576 vs. 32. One effect of this is time spread. (More frequency resolution requires a longer time window. This is a fundamental physical law.) For one or two passes, this is good, as the ear has masking in both the time and the frequency domains and L3 naturally exploits this additional dimension. The downside is that when many stages of L3 are used at low bit rates, the time spread can become audible (softening of transients and pre-echoes, mostly), and this is a bad thing. L2 does not have this problem. However because it is closer to the edge for S/N, multiple generations result in unmasking (noise and grit, mostly).

International Standards Organization (ISO)/MPEG proponents do not propose that a lot of passes of L3 be used. They advocate that L3 be used at ISDN bit rates for remotes and that L2 be used at higher bit rates in other parts of the signal chain. This is why ISO decided to recommend the layers as they did: L3 for 64 kbps/channel and L2 for equal to or greater than 120 kbps/mono channels.

My own experiments with codec cascading confirm that this is the right approach—the two coding methods seem to complement each other. Two passes of L3 sound noticeably better than two of L2; a pass of L3 followed by a pass of L2 also sounds better than two of L2.

DIAL-UP REMOTES ON POTS LINES

We can not always have an ISDN line at a remote site. Sometimes they are not available from the telco, or we just prefer not to use them because of the cost or the delay and trouble of getting one installed. Since cheap POTS dial-up lines are everywhere it makes sense to find ways to use them for program remotes. The problem is that the 300–3.4 kHz frequency response and limited dynamic range that the dial network provides are not generally adequate for modern broadcast needs.

Modem + Coding = Broadcast Audio on POTS Lines

A method that has emerged in the past few years is the so-called POTS codec. This piece of gear combines a high power coding algorithm with a fast modem. Of course fast is relative here. Recall that ISDN can supply a minimum 56 kbps bit rate, while the fastest modem is limited to 33.3 kbps—and very rarely achieves this speed, usually settling at around 24–26 kbps. (56 kbps modems have the fast rate only in the downstream direction; the upstream remains limited to 33.3 kbps.) Because our goal is to achieve something approaching broadcast quality, this is a very challenging bit rate for audio coding technology.

Generally, we employ a kind of coding that is optimized for speech and very low bit rates. The most common are taken from the code excited linear prediction (CELP) family, which have better audio quality for speech at very low bit rates than the MPEG perceptual coders used in ISDN equipment. They also have much lower delay, a critical characteristic when live interaction is required. Perceptual coders work by using an understanding of how the human ear works, while CELP algorithms model how the mouth produces voice sounds. Not surprisingly, CELP coders do a fairly poor job with musical signals. There can also be problems with background noises, such as applause.

Some equipment has the possibility to select from either CELP or an MPEG algorithm so that the user can decide which trade-off to make.

Frequency Shifting

Now that the digital systems previously described are available, the frequency shifting technique is used much less often. Nevertheless, it has its place, and many units are still in service. Frequency shifting offers a way to squeeze more high frequencies into a line than it will normally pass. More accurately, the process allows different frequencies than the usual 300–3.4 kHz to be passed through a POTS phone line.

Frequency-shifting units using a single phone line move all frequencies up by 250 Hz at the encode side and down by 250 Hz at the decoder as illustrated in Figure 3.10-21. The result is a 250 Hz improvement at the low end at the cost of a 250 Hz loss at the high end. This means a typical phone line's response will be changed after the shifting process to 50–3150 Hz. The 250 Hz loss at the top is not very significant due to the logarithmic nature of audio perception.

The shifting function is accomplished by heterodyning the input audio with a low-frequency carrier. The phasing single sideband (SSB) generation method is employed to allow only one sideband to emerge at the output—the carrier and other sideband having been cancelled in the SSB process. Encoding and decoding can easily be accomplished in the same unit, since only a simple signal path change is required for an encoder to decode, and vice versa. (See Figure 3.10-22.)

Subjectively, the resulting frequency-shifter-processed audio sounds less telephone like. However, the result of improvement at the low end without high end enhancement is often a somewhat muddy or flat sound. You can sometimes improve subjective quality by boosting the high end with a sharp EQ rise above 2 kHz. A parametric EQ or custom filter is preferred so that a high-Q curve can be obtained.

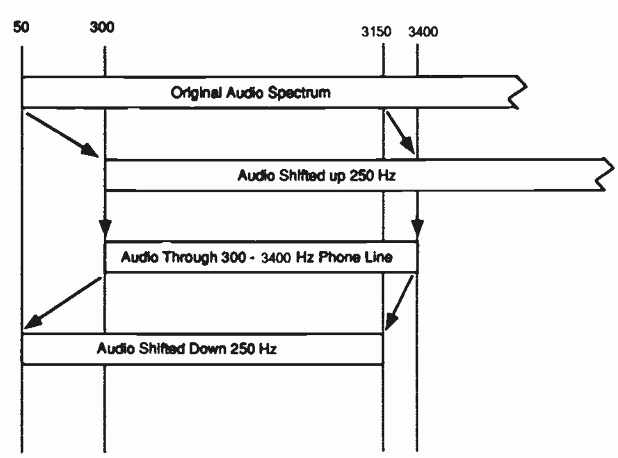


Figure 3.10-21. A frequency shifting bandwidth extender allows improved low end response at the expense of a small loss of high-frequency audio.

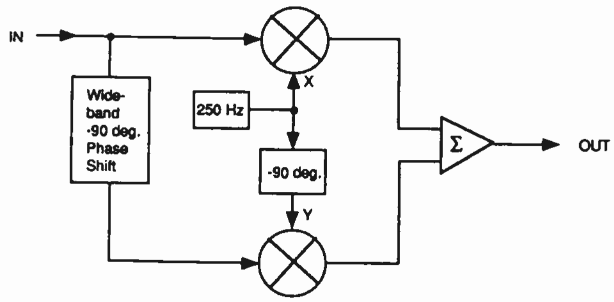


Figure 3.10-22. Single-line frequency extender uses SSB techniques to shift the audio 250 Hz at each end. A decode system is shown; signals at X and Y are reversed for encoding.

INTERFACING PRODUCTION INTERCOM SYSTEMS

To aid communication with the field crew during remote broadcast projects, connecting the production intercom system to dial-up telephone lines is often required. Smooth integration of live news remote feeds, for instance, requires that production personnel at all locations be able to communicate with each other in a simple, trouble-free fashion. This is especially true when multiple remote sites are involved, as for election coverage, major sporting events and telethons. Ideally, crews at each location would use the intercom system without regard for the distances involved. Most often, access to the dial-up phone network is available by wire or cellular, so an interconnection of the intercom system to the telephone network may be the solution.

4-Wire Intercom Systems

Four-wire systems are those in which the two speech directions are kept separated in the switching and distribution process. While it would be possible to use special four-wire telco circuits (or two standard loops) to maintain independent signal paths to remote sites, it is more economical and convenient to be able to use a single phone line. To accomplish this, we must create effective conversion between the 2-wire phone line and a port on the 4-wire intercom matrix. It will be necessary to separate the intermingled send and receive speech signals on the phone line with a 2-to-4-wire converter, or hybrid. One approach is given in Figure 3.10-23.

Transhybrid loss performance will be important when intercom stations with open loudspeakers and mics are to be used and when conferencing of multiple telephone lines is desired. In the first case, the acoustic coupling between the speaker and mic completes a feedback path which includes the hybrid. Clearly, the better the hybrid's isolation, the higher the feedback margin. In the second case, a feedback path exists from each active hybrid through all of the others that are conferenced to it. When the total gain exceeds unity, feedback results. The goal is to have the best possible transhybrid loss so that the maximum line-to-line gain may be achieved.

An auto-answer and disconnect function may be required for unattended operation. This circuit responds to a phone line ringing signal by activating the hybrid and de-activates the hybrid when the calling party hangs up. As discussed in the section on calling party control (CPC) a dial tone detector may be necessary to ensure reliable operation. The tone detectors are connected so as to respond to signals on the hybrid's separated telco receive audio signal. Were this not the case, and the detector was merely connected across the phone line, there would be a major problem when multiple lines are used together in a conference. Why? Because the tones would be conveyed to each line in use (through the intercom switching matrix) from every other line, causing all of

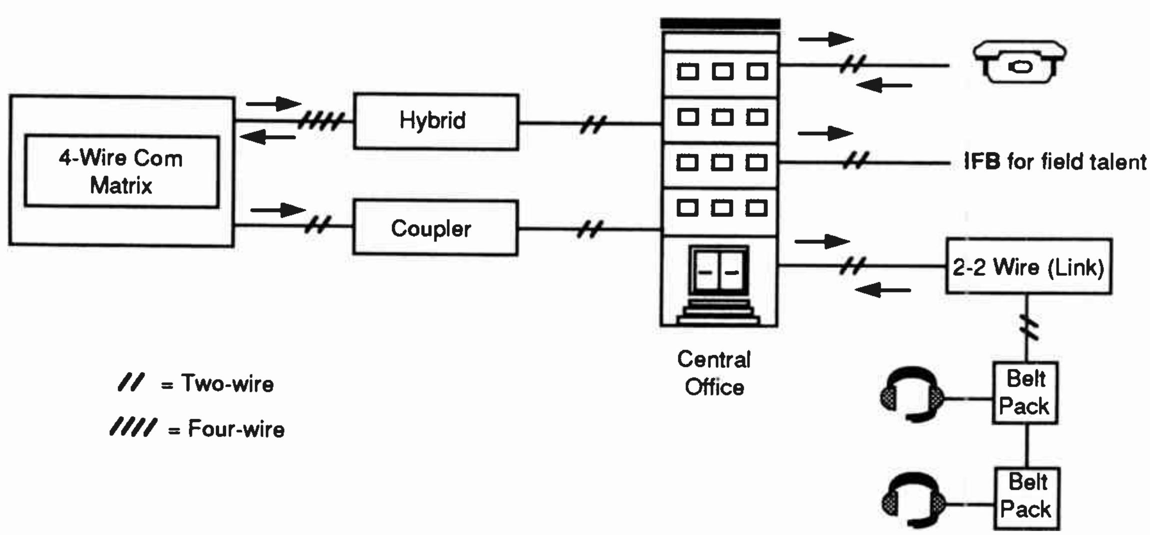


Figure 3.10-23. An arrangement which integrates a four-wire switching matrix with telco lines, an interruptible fold-back (IFB) feed, and a two-wire party-line intercom system for field production work.

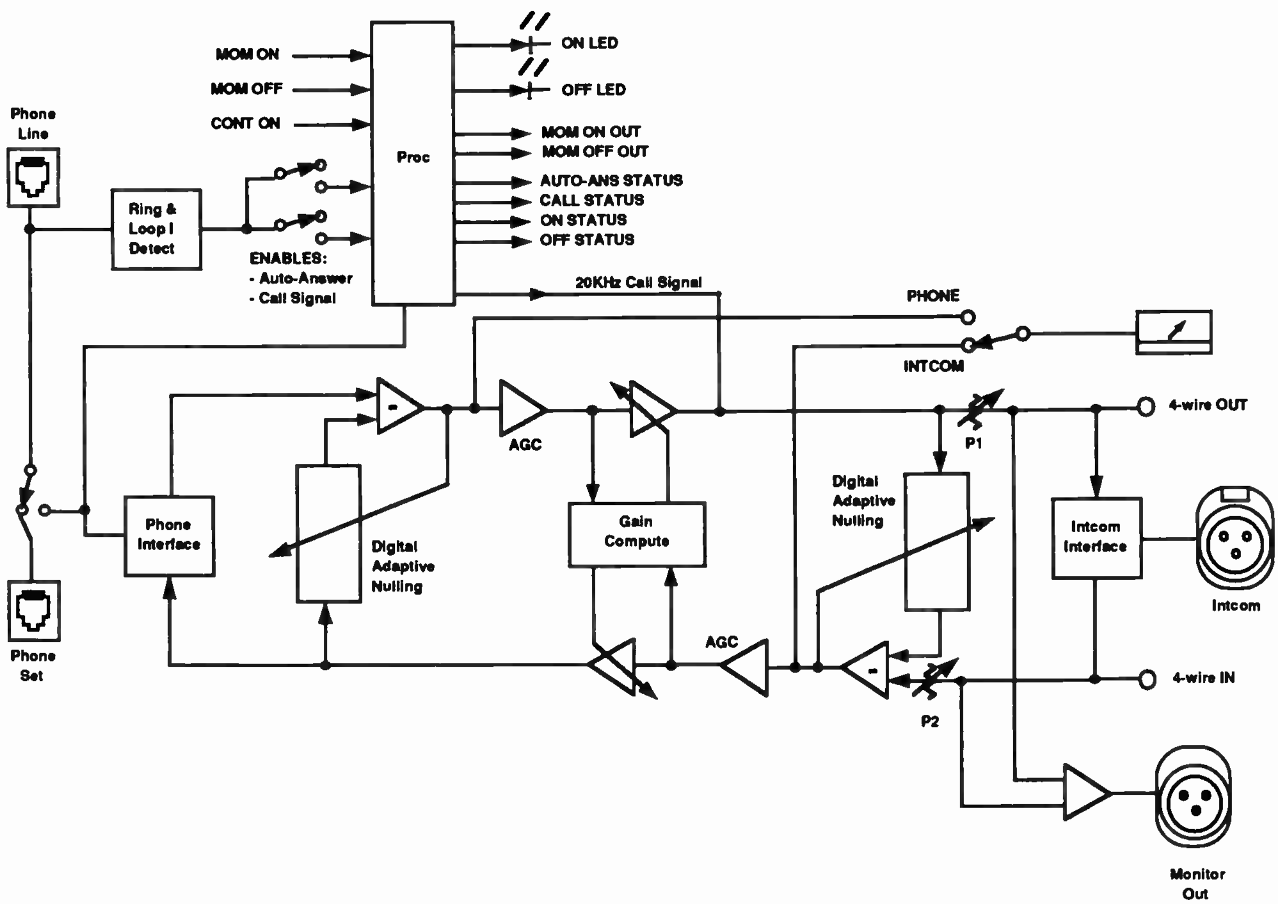


Figure 3.10-24. Two-wire intercom-to-telephone line interface (Telos "Link").

the detectors to respond to the tones from all of the other lines as well as its own. When one line's interface gets a disconnect, all of the others would turn off as well! Therefore, there is a critical requirement in this setup that transhybrid loss must be sufficient in order to be certain that any cross coupling is below the threshold of the tone detectors. The same situation applies with any DTMF detection that is used on a per-line basis.

2-Wire Systems

These are the popular party-line systems. Here, the interface to requires two hybrids. The hybrids are connected back-to-back so that the intercom hybrid's receive output is fed to the phone line hybrid's send input and vice versa. Appropriate gain and processing stages are inserted in the 4-wire path. This system is what telephone engineers call a 2-wire-to-2-wire repeater.

High quality hybrids are required to prevent feedback. As should be evident from Figure 3.10-24, the signals can feed around the loop and feedback could build up. This happens when the combined transhybrid loss of the hybrids is not at least as great as the gain in the two amplifiers.

As telephone circuits have widely varying and unpredictable end-to-end transmission characteristics, interfacing intercom systems to phone lines without gain and without AGC is not likely to work very well.

ISDN

Because ISDN circuits are inherently 4-wire, they are perfect for the intercom application. Used with a 4-wire intercom system, speech paths may be kept separated end-to-end. Applied to a 2-wire intercom system, the problem of maintaining sufficient hybrid balance is eased. ISDN lines are cheaper and easier to get than the special 4-wire lines sometimes used for intercom interconnection. Yet another benefit is that ISDN offers two channels so that production and talent feeds may be kept separate. Finally, a low delay coding method such as G.722 can be used to improve audio bandwidth in order to correspond more closely to the fidelity users are accustomed to on local links.

REFERENCES

Fike, John L., and George E. Friend, *Understanding Telephone Electronics,* Howard W. Sams & Co., 1984.

Flanagan, William A., *The Guide to T-1 Networking,* Telecom Library, Gilroy, CA, 1990. A superbly written and very complete description of the technology and use of T-1 service. Highly recommended if you need to learn about T-1. Available directly from Telecom Library at (800) LIBRARY. Incidentally, the Telecom Library is a very useful source of all kinds of material on telephones, from tutorials on 1A2 key systems to the latest on ISDN and digital technology. Call for their free catalog.

Newton, Harry, *Telecom Dictionary,* Telecom Library, Gilroy, CA,. Available directly from Telecom Library by calling (800) LIBRARY.

Teleconnect. Call (888) 824-9795 for subscription information to this monthly magazine. Edited by Harry Newton.

Telos Systems' website: www.telos-systems.com.

3.11
COMMON CARRIER AUDIO PROGRAM SERVICES

SKIP PIZZI
MICROSOFT CORPORATION, REDMOND, WA

INTRODUCTION

Broadcasters and the telephone company (telco) have been allied since the earliest days of both industries and will likely remain so. This chapter considers the opportunities for audio signal interconnections afforded to broadcasters by telcos, and examines how broadcasters can implement these hookups. At this writing, the telephone industry has nearly completed its quiet revolution from analog to digital distribution, so emphasis will be placed on the latter.

For clarity, the term circuit will be used to designate a telco audio path, supplanting the other commonly heard (and potentially confusing) synonyms of *loop, private line, leased line, program circuit,* and the like. Likewise, telco will be used to refer generically to all telephone companies, local, and long distance. LDS (long distance service) will be used to refer generically to all common carriers providing service between local telcos. General comments regarding telco service in this chapter will be limited to circuits used specifically for audio program transmission. For interfacing audio to/from the standard analog dial-up network (*Plain Old Telephone Service,* or POTS), see Chapter 3.10.

Dealing With Telco

The best way to minimize problems with your telco is to establish a good working relationship with the appropriate personnel, and to understand as much as possible about the company's service and operation. If the station's staff comes across as friendly and knowledgeable, but also professionally firm and businesslike, things should go well. If possible, keep the station's liaison to the telco limited to one staff member, and try to always deal with the same person at telco as well.

For ordering digital circuits, a recently instituted option is the data-line brokerage service. Here, the station gives its time, place, and quality of service requirements to a third party, who books the line for the station at no charge. (Like travel agents, these services receive commissions from the telcos whose circuits they book.) This service can be especially helpful in long distance applications, where two local telcos (one foreign to the station) and an LDS are involved. Following the travel-agent model, data line brokers may also be able to book the service at a lower cost. In some cases, any extra customer-provided equipment (CPE) required for the circuit may be rented from these brokers, as well.

ANALOG CIRCUITS

Analog audio circuits are still offered by some telcos under the schedule shown in Table 3.11-1, but in major metropolitan areas there are very few new lines being installed. Telcos are instead replacing the existing circuits that remain with digital equivalents. Meanwhile, LDSs have already phased out such analog services for their long-haul paths.

Where they are still offered, costs for both service and installation of analog program circuits continue to increase. (The exception is in some rural telephone service areas, where alternative digital service is not yet a viable option.) Installation of an analog program circuit generally requires several hours of an experienced technician's work to equalize the line, as opposed to digital services, which take much less time and trouble to pass spec at installation.

Past practice with local analog circuits often provided wider bandwidth than what was ordered, whereas digital services typically cut off exactly as specified. Increasingly, though, analog services that have to pass between telco central offices (COs) or switching centers will make the trip bundled on an interoffice digital carrier anyway, whereupon any excess bandwidth will likely be removed.

Obtaining and Testing Analog Telco Circuits

As Table 3.11-1 explains, analog circuits are available in a variety of bandwidths, and under temporary or permanent status. Check with your local telco to see which, if any, remain available and what they cost. Installation charges for permanent lines may be much higher than those for temporary service, because the

Table 3.11-1
Classes of service for analog telco audio circuits.

CLASS OF SERVICE	APPROX. BANDWIDTH	FULL- OR PART-TIME
Type 6002	200–3,500Hz	PT
Type 6003	200–3,500Hz	FT
Type 6004	100–5,000Hz	PT
Type 6005	100–5,000Hz	FT
Type 6006	50–8,000Hz	PT
Type 6007	50–8,000Hz	FT
Type 6008	50–15,000Hz	PT
Type 6009	50–15,000Hz	FT

telco may want to actually install new wiring rather than permanently occupy any pairs on its existing network cables. In an increasing number of areas, analog audio circuits will not be offered at all, in which case the reader may skip the remainder of this section. For those legacy users of analog audio program circuits, however, the following is provided.

Type 6008 and 6009 (15 kHz circuits) service may be ordered as a stereo pair, incurring a one time installation surcharge for *stereo conditioning*. (Some telcos also offer it for 6007/8 [8 kHz] service.) This ensures that both lines are routed together throughout their runs, so that interchannel phase differences will be minimized. A third line can be ordered for backup, and this too should be included in the stereo conditioning. Although billed routing (as the crow flies) may be a short distance, actual routing of the circuits may be much longer and indirect, providing ample opportunity for phase differences to occur. (Approximately 5 μsec time difference occurs for every mile of path length difference.)

Lines should be ordered well in advance of your need for them. Check with your telco for their preferred lead time. Always specify a start date at least one business day earlier than the actual requirement, to allow time for your own tests on the lines to be performed. Check frequency response, signal-to-noise ratio (S/N), distortion and headroom. For stereo pairs, check relative phase response and relative polarity.

Frequency response should be at least within ±3 dB of what was ordered. Be sure to check outside the passband, because response may not roll off but instead rise beyond the cutoff frequency. On occasion, in order to get a line to meet specifications, telco equalizers may be used to boost the extreme frequencies, and if done improperly, the response may indeed be flat to the cutoff frequency, but then increase for another octave or more before finally rolling off. This will result in audible consequences from the altered response and reduced headroom, particularly if a noise reduction system that preemphasizes high frequencies is used on the circuit.

(Use caution with any complementary analog noise reduction (NR) system used on a telco circuit. As a general rule, be sure that the audio fed to the NR encoder is pre-filtered to match the line's response, so that the audio passband seen by the encoder is relatively similar to that seen be the decoder after the audio has passed through the circuit. If the decoder sees a substantially narrower bandwidth than the encoder, it will not decode the noise reduction process in a complementary fashion.)

The proper procedure for calibrating levels on analog program circuits employs sinewave test tones of 400 Hz or lower, fed at the telco program operating level (POL) of +8 dBm. frequencies above 400 Hz must be sent at the telco test level of 0 dBm to minimize crosstalk into other circuits via capacitive coupling of higher frequencies. To keep things simple, you may want to run all frequency response tests at 0 dBm.

For measuring noise, telcos use a slightly unorthodox approach. They consider a noise level of −90 dBm to be *absolute quiet* or noise free, and measure noise from that reference point. The unit used is dBRN (RN for reference noise). Therefore, a −50 dBm noise level would be called 40 dBRN by telco. If the telco's specified audio reference level of +8 dBm is used by the customer on this circuit, a 58 dB SNR is achieved in this case.

The greater the distance an analog circuit travels, the noisier it becomes. The wider a circuit's bandwidth, the quieter it needs to be. Although specs vary between telcos, noise specifications generally reflect those observations. A typical noise level for local 15k Hz circuits is 33 dBRN or lower, providing 65 dB or better SNR. Long distance analog circuit noise levels generally hit 40 dBRN (58 dB SNR) for 15 kHz circuits. Noise levels for a 5 kHz service are around 46 dBRN (52 dB SNR) for local and 56 dBRN (42 dB SNR) for long distance service. Again check with your local telco for its specifications, and always verify that they are met.

Some telcos offer a lossless or zero loss option, in which the circuit acts as a unity gain device. This is contrasted to a standard circuit, which may exhibit up to 30 dB of loss. (Audio signals in copper cable drop at about 1 dB/mi from broadband resistive losses. Frequency selective reactive losses occur at greater rates, but these are remedied by receive end equalization.) The additional telco amplifiers required by lossless lines may have a detrimental effect on the distortion and headroom performance, and may not do much to reduce the overall static noise floor, but they are useful in densely trafficked urban areas where crosstalk and impulse noise is prevalent.

Static or random noise is far less objectionable than those coherent noises sometimes found on telco circuits. Such noises can be caused by capacitive coupling between adjacent pairs in multipair cable, dial pulses and other switching, inadequate common mode rejection, and carrier beating (causing high pitched sings or tones). Because of this, circuits should be carefully auditioned at the receive-end for a while after installation, without any audio on the line. Check circuits again prior to each on air use. Report any crosstalk or impulse noise problems to telco at the first sign of trouble.

Although there are usually no published specs for distortion or phase response on telco circuits, total harmonic distortion (THD) should be <0.25% on 15 kHz lines. For stereo pairs, relative phase response should be within 30° across the passband. Widely divergent frequency response between the two circuits in a stereo pair is a tip off to check phase response carefully. When a spare third circuit is ordered, the two closest to each other in frequency and phase response should be designated as the main pair, assuming distortion and noise are equivalent across the three.

The maximum level guaranteed on telco circuits is +18 dBm, which implies that only 10 dB of headroom will exist above the +8 dBm reference level. A sensible alternative is to use +4 dBm as a reference level (most professional audio hardware uses this level anyway), thus allowing a more realistic 14 dB of headroom, at the expense of 4 dB less SNR—generally a worthwhile tradeoff.

Interfacing Procedures

Figures 3.11-1 through 5 illustrate some do's and don'ts of audio interfacing to analog telco circuits. This is one area of today's audio where 600 Ω impedance matching is still important. Telco equalizes its circuits for flat response under the conditions of a 600 Ω source impedance and 600 Ω termination. Because of the reactive components in long wired paths (their transmission line behavior implies complex impedances), varying these impedances will affect the frequency response of audio carried on the circuit. Most contemporary audio equipment expects a bridging interface arrangement, so output source impedances are typically much lower than 600 Ω, and input impedances

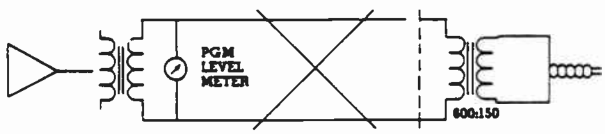


Figure 3.11-1. DON'T meter the signal across the input to the repeat coil (transformer at right) and DON'T feed the repeat coil directly from another transformer, if possible.

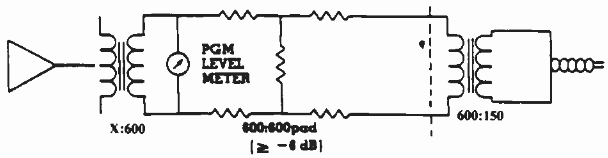


Figure 3.11-2. DO place a pad between the output device's transformer and the repeat coil; and DO place a level meter before the pad, calibrated for the voltage across the pad's output when terminated with a 600 Ω resistor.

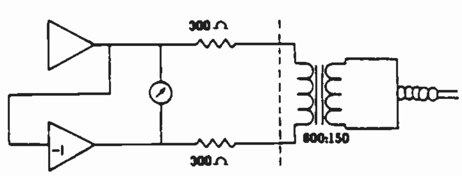


Figure 3.11-3. DO feed low-impedance sources (e.g. op-amps) through a 600 Ω differential balanced pad; and DO place the level meter before the pad, calibrated as Figure 3.11-2.

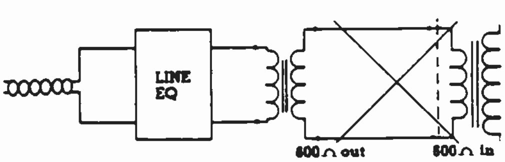


Figure 3.11-4. DON'T terminate the receive-end repeat coil (transformer at right) with another transformer, if possible. Its loading may vary with frequency from the true 600 Ω resistive termination used in line-up causing level and frequency response variations.

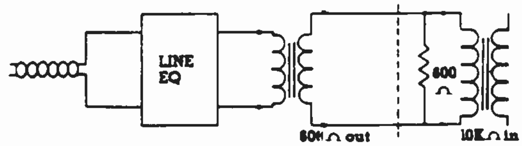


Figure 3.11-5. DO terminate the receive-end of a program circuit with a 600 Ω resistive load and bridge the load with a high-impedance transformer or active balanced input. Note: Common mode rejection of the transformer or active input must be considered if the distance from the repeat coil is great or is near other lines which induce cross-talk. Ideally, the secondary of the repeat coil should be resistively terminated. An active, balanced input circuit does this nicely, provided any RF is by-passed before the first stage of amplification. An alternative is the use of an input transformer with a high-impedance, bridging input. This allows the 600 Ω resistor, as show in Figure 3.11-5 to match the repeat coil. Another version, often seen on the input of broadcast line amps and modulation limiters is a 600 Ω h-pad; effectively the reverse of Figure 3.11-2. Following the steps shown in Figures 3.11-1–3.11-5 usually insures that transmission loss and frequency response closely match those of the phone company set up.

are much higher. Yet the telco does not guarantee flat response without 600 Ω matching conditions.

If a typical low impedance mixing console output is used to feed a telco circuit directly, it will generally cause a rising high frequency response to be received. This is not telco's responsibility, but rather the customer's. It is therefore essential that the device directly feeding the line have a 600 Ω source impedance, and be capable of driving such a balanced load to at least +18 dBm across the passband. Many mixers—even expensive ones—are not designed to do this, so an appropriate outboard line amp or distribution amplifier (DA) fed by the mixer should be used.

A DA will also provide isolation between the mixer and any other inputs, and more importantly, between all these devices and the telco circuit. Drive the telco circuit with its own DA output—no other inputs (even bridging ones) should be paralleled to it. DC voltages may appear on telco circuits, so this isolation is essential. For the same reason, a transformer coupled DA is the better choice over a transformerless design in this application. High quality transformers can be quite helpful on remotes, especially when interfacing with a telco. On most 15 kHz circuits (and some others), though, telco repeat coils appear at each end, providing this isolation for you. In this case, actively balanced outputs can be used safely, but flat frequency response is only guaranteed if they are interfaced to the repeat coil in the manner shown in Figure 3.11-3. A caveat here: remember that the active DA still has to drive a 600 Ω balanced load to +18 dBm across the passband, and many popular operational amplifiers do not have the current-drive capability to do so. To avoid crosstalk to other circuits, do not conduct lengthy tests at +18 dBm.

The "H" pad in Figure 3.11-2 does not provide a 600 Ω source to the line by itself. If the secondary of the output transformer on the left of the diagram were not 600 Ω, the pad would not set things right. The pad is there to provide isolation of the meter or a test oscillator from the repeat coil.

A simple test to determine the actual source impedance of a device's output is as follows. Place a bridging input voltmeter across the device's output while feeding 1 kHz at reference level. (No other inputs should be connected.) Terminate the output with various resistances, and watch the meter. The device's source impedance is the value of the resistor that drops the level by 6.02 dB. Then verify that this level drop is consistent at other frequencies.

Note also that the meter on the transmit end shown in Figures 3.11-2 and 3.11-3 should be used only for initial absolute level calibration, and not for relative levels in frequency response tests. For verification that consistent level is being transmitted at each frequency during response tests, and for realistic end-to-end results, put the oscillator further upstream, and use a more isolated meter for reference. This typically is done by feeding the oscillator into a properly telco interfaced mixer or DA input (or using the mixer's inboard multifrequency oscillator, if it has one), and recalibrating for oscillator drift using the meter on the oscillator, mixer, or DA, not the downstream meter shown in the figures.

Failure to follow the procedures outlined above will result in poor frequency response, distortion, calls from the telco warning of too high a level, too much noise and general unhappiness with the service.

DIGITAL AUDIO CIRCUITS

As with much of the technological progress in our industry, the digital audio revolution that is now taking place in the broadcast world began at the telco. The transmission of data is nothing new to telco but the high data rates required for digital audio transmissions had previously rendered the availability and cost of such service out of the practical range. Data compression (or bit-rate reduction) systems have made possible broadcast applications of data transmission paths that previously were only useful for computer interconnection. Reductions from earlier data rates for digital audio transmissions of 4:1 or higher have become commonplace and compression ratios of 12:1 without audible penalty are now emerging.

This implies that audio program circuits are ordered from telcos not as audio services per se, but simply as digital circuits with appropriate bandwidth for the transmission task at hand. The actual signal(s) passing through the line are of little consequence—a digital circuit cares not whether the program it carries is a symphony or a spreadsheet.

Data Compression

Although these data compression algorithms are viewed today as major breakthroughs, history will likely look upon them as natural evolutions, and consider the earlier linear pulse code modulation (PCM) systems as dinosaurs. While the straightforward nature of linear PCM may have been helpful in making the transition from analog systems, especially where bandwidth was cheap and available, it is an inefficient method for encoding digital audio. The resolutions of today's linear PCM systems are often overkill in terms of the actual needs of most listeners, and significant reductions in actual transmitted data can be achieved by applying data compression algorithms to the datastreams that linear PCM conversion produces. At present, linear PCM of as high a resolution as economically feasible is still a good idea for the original conversion of analog signals to the digital domain, and for any digital production or signal processing. But for signal delivery systems (and in some cases, for long-term storage), data compression is an appropriate tool.

Earlier compression systems (again pioneered by telcos) used a purely statistical or numerical analysis of the datastream's coding redundancies. These lossless systems could be used to reduce the data rate of any kind of data transmission. More recent lossy systems are designed for exclusive use on audio data. These coders acknowledge the limits of the human listener's sensory perception and exploit them for much higher compression ratios. Such perceptual coders are therefore based on psychoacoustic models, and owe their coding efficiencies to an appreciation of the audience's tolerances.

Data Rates

Like their analog counterparts, digital circuits come in various bandwidths. But rather than specifying cutoff frequencies of the audio passband, telcos specify digital circuits in terms of their data rates. Broadcasters now need to consider bandwidth requirements in two dimensions: that of the audio signal itself and the data-transmission path. The former is specified directly (in Hz), while the latter is specified as a data rate (in bps). The resultant audio quality through a digital transmission path is a function of the bandwidth of the digital circuit and the performance of the coding algorithm at that data rate. For example, linear PCM will provide only POTS-like results at 64 kb/s, but ISO/MPEG-1 Layer III Audio Coding can provide near CD-quality mono audio on the same 64 kb/s circuit.

Therefore, broadcasters must develop an acuity for the appraisal of various data rates capabilities, both with and without data compression. In the linear PCM (uncompressed) mode, the data rate requirement of a given audio signal is determined by simply multiplying its sampling frequency (in Hz) by its resolution (in bits/sample). For example, CD quality audio uses 44.1 kHz sampling at 16 bits/sample resolution, requiring a 705.6 kb/s data rate, per channel (stereo requires doubling that data rate to 1.411 Mb/s), before adding any error correction overhead. A digital audio compression algorithm capable of 4:1 data rate reduction can reduce that signal's resolution to an average of 4 bits/sample (while leaving its sampling frequency alone), therefore providing a 176 kb/s data rate for mono audio. Table 3.11-2 shows some other data compression ratios for audio, and their resultant data rates, at several common sampling frequencies.

Table 3.11-2

Data compression table showing a range of compression ratios and their resultant output data rates at a variety of sampling rates (fs). Audio bandwidth is approximately one half of fs. Data rates shown are for a single audio channel (mono).

RESOLUTION (avg. bits/sample)	COMP RATIO	OUTPUT DATA RATES (kbit/s) fs = 48 kHz	fs = 44.1 kHz	fs = 32 kHz
16	1:1	768	706	512
4	4:1	192	176	128
3	5.3:1	144	132	96
2.67	6:1	128	118	85
2	8:1	96	88	64
1.45	11:1	70	64	46
1.33	12:1	64	59	43

Because the sampling rate is not changed by a data compression system, frequency response and time domain performance retain the same specifications typical of most linear PCM conversions. Processing delay is introduced by these codecs, however, and its length is generally in direct proportion to the amount of data compression applied.

Telephone company installations and tariffs provide a variety of services, with more new services continually being deployed in many cities. Table 3.11-3 shows some current services and their data rates.

One of the most important differences to the broadcaster between analog and digital circuits is that most digital services are provided bidirectionally. This fact should not go unnoticed when making cost comparisons. Although interfacing hardware for return path channels must still be provided, their circuits require no separate costs or orders.

Note also that telcos have always had their own insider vocabulary and set of acronyms, but since the introduction of digital services, this has expanded dramatically. Many of these are explained in the *Glossary of Digital Telco Terms* at the end of this chapter.

For overseas links, rough equivalents to each of the domestic services shown in Table 3.11-3 do exist outside the U.S., but their actual data rates differ. Format conversions are therefore required for international transmissions, but most LDSs can handle this for the broadcaster. For example, European telcos offer an

Table 3.11-3

Some current U.S. digital telephone data services and their data rates. The ISDN data rates do not include the 16 kbps D channel (for BRI) or the 64 kbps D channel (for PRI), nor do they include framing, synchronization and other overhead bits.

SERVICE	DATA RATE (bits/sec)	MODE
Switched 56	56 k	Switched
DS 0	64 k	Dedicated
ISDN-BRI	128 k	Switched
ISDN-PRI	1.472 M	Switched
DS 1 (T-1)	1.544 M	Dedicated
DS 2	6.312 M	Dedicated
DS 3	44.736 M	Dedicated

E-1 rate of 2.0 Mb/s, in lieu of the North American T-1 rate of 1.5 Mb/s.

DS1 or T1 Service

Digital audio transmission on DS1 (or T1) lines has become widely available, and is often cheaper than standard analog circuits in both service and installation charges. (See *Glossary* for distinction between DS1 and T1 nomenclature.) DS1 is a bidirectional 1.544 Mb/s serial data link. The data rate calculations above show how DS1 can carry a single, linear PCM stereo audio signal, or several such compressed channels.

DS1 service is extremely reliable. Its bit error rate (BER) of 10^{-9} (the probability of error reflected by the specification of no more than one erroneous bit in 10^{-9} transmitted) is among the lowest available. By way of reference, IEEE and ITU-T have both established 106 as the BER required for data customer satisfaction.

The data carried on a DS1 circuit is actually a multiplex of 24 data channels, or slots, of 64 kb/s each. (An additional 8 kb/s is reserved for sync data.) These individual 64 kb/s slots are called DS0 channels. For standard telco T carrier use, each DS0 carries a digital voice grade circuit, using the so called μ Law (nonlinear 8 bit) algorithm on 8 kHz sampled audio. When a customer leases a DS1 circuit, it can be configured to carry any bandwidth channel that DS1 hardware is available for (3.5 kHz, 5 kHz, 7.5 kHz, 15 kHz) in any combination, up to the customer's payload data limit. When a customer leases a full DS1, the telco may use one DS0 slot for framing and other overhead, in addition to the 8 kb/s synchronization slot, leaving around 1.4 Mb/s for customer data. Check with your telco for its exact rate.

A rack of coding and multiplexing hardware appears on each end of the DS1 line, usually as customer provided equipment (CPE), and the circuit can be reconfigured simply by changing the appropriate cards in the proper slots in the racks at both ends. The customer can perform these reconfigurations at any time, without telco involvement or notification.

Unlike the labor intensive installation and equalization of an analog circuit, putting in a T1 circuit has become nearly as routine as a standard dial up telephone service installation. This and the capacity gluts in some areas continue to lower costs for DS1 service. Customers' use of digital compression systems on DS1 channels will only increase this economy. Whereas a 15 kHz audio channel had in the past typically required six DS0 slots, current hardware implementing perceptual coding reduces this to two (or even a single) DS0s.

Fractional T1

In some areas, Fractional T1 service is available, generally for intraLATA (local) applications only. This service allows a customer to lease only the number of DS0 slots on a DS1 circuit that are needed for a particular application. Although installation charges will be about the same as a full DS1, service charges may be substantially reduced for many remote audio applications.

Switched 56

Another telco digital offering that has been used by broadcasters is the Switched 56 service. This facility has been available for a number of years for local service in many metropolitan areas in the U.S., and from some long distance carriers. It provides a bidirectional 56 kb/s data path for use with dial up terminals. Like POTS, a monthly service/network access fee is charged (which often includes at least some free local calling), with long distance calls billed by the minute, at rates similar to regular dial up long distance service. In some cases, local calls are also billed for connect time. An installation fee is also typically charged for initiation of service.

A switched channel service unit (CSU)—the equivalent of a telephone instrument and data interface—is provided by the telco, or may be purchased by the customer. It allows voice or data interconnection, and dial up routing of the data path to any other similarly equipped destination on the network. For broadcast use, additional codecs are required, as shown in Figure 3.11-6, to feed wideband audio. These codecs are not available from telcos, and must be purchased by the customer. The typical codec used for Switched 56 service is ITU-T G.722, which provides 7.5 kHz mono audio (see Figure 3.11-6).

An obvious savings is possible with such a switched approach, since dedicated circuits need not be installed between a radio station and all of its remote sites. The station and each site need only to be wired for the switched service (with each line going to the telco CO only, and not end to end) and the station can dial up any remote site as needed. The station may require two or more lines for accessing multiple remote sites simultaneously or in quick succession.

CSUs are available in two wire or four wire versions, with two wire types costing less. The choice between two and four wire operation is not up to the customer, however, but to the local telco serving the area.

Switched 56 is available to Europe and Japan, as well, although it is a 64 kb/s service in most overseas locations. Conversion hardware is available for such an application. Like the higher speed channels mentioned above, no international standards exist for these switched services today, although they are currently under development.

Some telcos offer a similar unswitched service, in which a single DS0 channel can be leased on a monthly basis. For heavy point to point users, this may be more economical than a switched approach. As single DS0s, these circuits operate at 64 kb/s, and their terminal hardware is less expensive because it need not accommodate switch signalling.

New installations of Switched 56 service are now rare, with most telcos (and users) preferring the more recent Integrated Services Digital Network (ISDN).

ISDN Service

At this writing, ISDN is available in most urban centers, and will soon be universally deployed in the U.S. The most common variety of ISDN service is the Basic Rate Interface (BRI), which provides two 64 kb/s paths (bearer or "B"-channels) and one 16 kb/s circuit (data, delta or "D"-channel). As a result, this service is also referred to as 2 B + D. Bearer channels carry customer data, while data channels carry signalling and call routing data. This feature is a significant departure from POTS and Switched 56 service, in which the call signalling data is routed via the same data path as the program audio or data. (This is why DTMF tones and call-waiting beeps or pulses are audible on a POTS call, for example.) Such interruptions can cause problems to data communications, which benefit from continuous, bidirectional connections. ISDN's use of such *out-of-band signalling* provides significant improvement in speed, throughput and robustness for switched data communications.

ISDN-BRI is intended as a standard residential service, while business service is provided by ISDN's *Primary Rate Interface* (PRI). This service provides twenty-three 64 kb/s B channels, and one 64 kb/s D channel (23 B + D). (The ISDN-PRI D-channel operates at a higher rate than the BRI service because it has more B-channels to manage.) The data payload capacity of the ISDN PRI service (1.472 Mbits/s) is roughly equivalent to the data payload capacity of a T-1 circuit (1.536 Mb/s).

While the above data rates are common for ISDN, it should be noted that some ISDN systems in the U.S. offer users only 56 kb/s of payload data on their B-channels.

ISDN is a bidirectional, customer switched service,

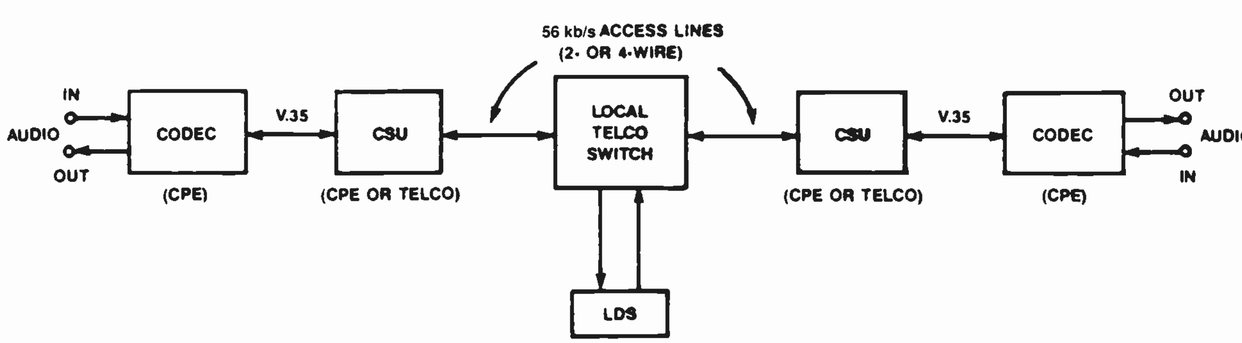


Figure 3.11-6. Block diagram of Switched 56 circuit path. (See Key Terms for acronym definitions.)

operating as a dial up, billed minute data network, allowing both circuit switched and packet switched operations. Its multichannel nature allows simultaneous voice and data to separate destinations (like multiline POTS service) or other applications in which multiple B-channels are combined to provide a single, higher-bandwidth call. This is done via a process called *inverse multiplex* (IMUX). Billing remains separate for each of the multiple lines in an IMUX, however.

The ISDN equivalent of the POTS telephone instrument is called the *terminal adapter* (TA), and it is available in several varieties, for the control of one or more BRIs or PRIs. Each TA is connected to the ISDN network via a standard *network termination* (NT), which varies with the type of terminal hardware and software installed by the local telco. In North America, all telcos have agreed to use a single standard interface, which requires the use of the NT-1 termination. While originally supplied as a separate unit for easier transport of a TA between service areas, the standardization has encouraged most manufacturers to include the NT-1 internally in most terminal equipment.

ISDN BRI has already become the delivery method of choice for many radio remotes, offering mono or stereo 15 kHz broadcast audio plus communications to/from a remote site. Often a single B-channel is used with one of several audio codecs capable of providing high-quality mono audio transport at a 56 or 64 kb/s rate. In other cases, two B-channels are IMUXed for high quality stereo at 112 or 128 kb/s. Numerous devices are available for such applications, and most include an integrated ISDN terminal adapter, NT-1 and codec, plus machine-control closures (see Figure 3.11-7). Most of these devices include codecs with variable data rates, and some may include multiple coding algorithms as well. Appropriate coding algorithms include ISO/MPEG-1 Audio Layers II and III, APT *x-100,* Dolby *AC-2,* Lucent PAC and MPEG *AAC.*

For more information on ISDN, including its use in voice transmission applications, see Chapter 3.10, "Telephone Network Interfacing." This chapter also contains information on the use of POTS codecs, which can provide moderate quality mono audio on standard dial-up phone lines.

DSL Service

An emerging technology that may play a role in future broadcast audio transport is the *Digital Subscriber Line* (DSL) service, which provides multimegabit connectivity over the telcos' existing twisted-pair infrastructures. It accomplishes this by implementing adaptive DSP at both ends of the signal path, thereby compensating for the distortion inherent in long copper lines. Path lengths for DSL are limited, but systems are designed to accommodate the typical telco customer-to-central-office runs of 12,000 to 18,000 ft.

DSL paths are dedicated, point-to-point services, and intermediate repeater amplification is not used. Two different transmission technologies are employed: *Discrete Multitone* (DMT, developed by Amati Communications), and *Carrierless Amplitude/Phase modulation* (CAP, from AT&T).

Several varieties of DSL are available. Of greatest interest to consumers is the *Asymmetrical Digital Subscriber Link* (ADSL), which offers a wider downstream bandwidth (from telco CO to customer) than its upstream capacity (from customer to CO). More important to broadcasters may be the *High-Speed Digital Subscriber Link* (HDSL), which provides equivalent capacity in both directions (up to 6 Mb/s). It may soon provide an economical alternative to T-1 service.

COMMUNICATION LINES

For remote broadcasts, circuits for transmission of program audio often must be complemented with separate communication interruptible fold-back (IFB) or talk-back lines. When analog program circuits are used, communications are usually carried on standard POTS lines. When digital services are used for program back-haul, their bidirectional nature usually allows monitoring and communications to be carried on the return side of the same lines. Nevertheless, it is still a good idea to have at least one POTS line on hand in such cases.

Communications lines may be used purely for talk-back if off-air monitoring is possible on site. Remember that the codecs used for digital services can add significant delay (on the order of 100 ms in each direction in some cases), as can studio-to-transmitter links and other components of the air chain. The cumulative delay from these devices can cause monitoring problems for talent listening to their own voices or in conversation with others. For this reason, dedicated (analog) communications lines may be useful even when air monitoring or digital return is possible.

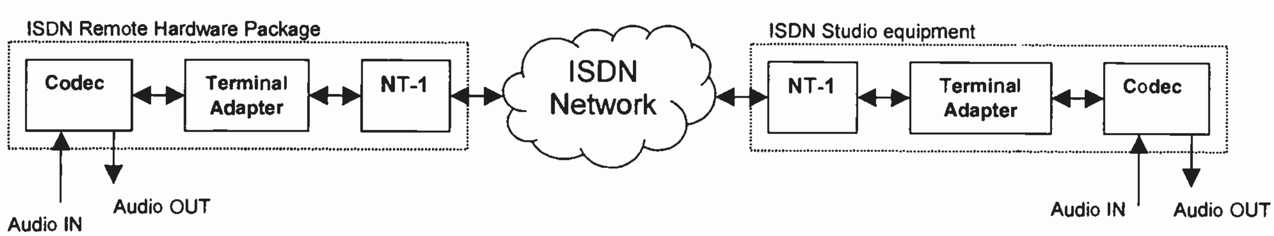


Figure 3.11-7. Block diagram of typical ISDN-BRI service used for a radio remote.

When using POTS communications lines, one or more standard dial up telephone circuits can be ordered for program audio monitoring and communications. Alternatively, an RF communication system (hand held radio or cellular phone) could be used. If wired telephone service is chosen, the station should provide its own instrument, which may be equipped with visual ring/bell cutoff and headset attachment.

There are also some variations to consider over a standard dial-up service. A communications phone line (comline) can be set up as an *off premises extension* (OPX) of the station's *private branch exchange* (PBX) system, or the line can be a dedicated *private line* (PL) that only runs between the station and remote site, in an unswitched form. In the latter case, such a PL is normally equipped with *automatic ringdown* (PLAR), in which special hotline type phones are installed by the telco, having no keypad or dial; the phone at one end rings whenever the phone at the other end is picked up. Services of this type (and their specific nomenclature) may vary among telcos. These services are also typically provided by a different department than the ones that handle program audio circuits or ISDN, so coordinate carefully. Typically, this part of the telco is the same one that handles all dial up installation and service calls from the general public, so lead time requirements and service call response time will often be longer than the program audio circuit or ISDN departments.

T-1 Applications

Using the return path on T-1 circuits for communications and backfeeding requires appropriate terminal hardware on each end. Channel configuration need not be the same as the circuit's other direction, and because narrowband (lower data rate) channels are usually all that are required, the number of communication channels on the return path can exceed the number of program audio channels on the transmit side, if necessary. Another possibility is the luxury of wideband backfeed and communications, such that the receive audio quality at the remote site is as good as the transmit quality.

For permanent interfacility hookups, a T-1 circuit could carry audio transmissions between sites plus telephone service, such that both locations can have their phones connected to the same PBX. If the locations are far apart, a *foreign exchange* (FX) arrangement can be established, whereby one facility can place calls to the other facility's telco service area without incurring toll charges. The PBX system is programmed to recognize area codes of the two cities involved (or special internal access codes), and it directs appropriate calls to FX lines on the T-1 rather than placing them as regular long distance calls.

COST OF SERVICES

For local service, there is normally only one provider of telco circuits, analog or digital, which is, of course, the local telephone operating company. So there is little competitive choice in the matter for intraLATA service pricing. Check with your local telco frequently, though, to see which services are more economical, for either permanent or temporary applications. Analog services continue to increase in price in most areas (if they are even available), while the cost of digital services declines at varying rates around the country. This makes it worthwhile to keep a close eye on these changing rates in your area.

For long distance paths, there is significant competition for interLATA digital service, so prices are kept low, and continue to drop. (There is little or no interLATA analog service available anymore.) Shop around for the best deal on interLATA digital service, or use a dataline broker. The list of common carriers in this business is long, and continues to grow. Secondary services may also be offered, such as switching and monitoring.

Back-up is also an important issue, particularly for live broadcasts. Many broadcasters establish primary and secondary backhaul and communications paths, using different services for each. (For example, ISDN primary and POTS secondary services.) This adds a level of redundancy because the services are typically routed differently. Costs of each service are an important factor, particularly when broadcasters are paying for something (the secondary service) that they hope they never have to use. The use of a lower-cost service for back-up therefore makes both technical and economic sense, and it is a highly recommended practice.

For those broadcasters who book a significant number of telco services, it also may be helpful to understand how telco billing works. Figure 3.11-8 shows the basic route-billing concepts followed by local and long distance common carriers for dedicated (non-switched) lines.

The importance to radio of interconnection, immediacy, and fidelity underscores the importance of a long term relationship between broadcasters and telcos. Good awareness of and rapport with telcos are essential to the daily work of the broadcaster. In today's context, that means keeping abreast with the changes in telecommunications that affect broadcasting as the digital revolution rolls along.

KEY TERMS

ADPCM. Adaptive Differential Pulse Code Modulation. A form of digital coding more efficient than linear PCM because it only codes the difference between one sample and the next, instead of assigning a fully discrete value to each sample. It also adapts its coding to the signal values currently under process. Considered a form of statistical data compression.

ADSL. Asymmetric Digital Subscriber Line. A high-speed digital service that uses telcos' existing copper (twisted-pair) infrastructure to provide data at a higher rate *to* the customer's premises than *from* it. For example, service may run 1.5 Mb/s from the telco CO to the customer, but only at 64 kb/s from customer to telco.

LEGEND

CM — Channel Mileage (per air-mile billing basis)
CO — Central Office (LEC switching center: also called SWC—Serving Wire Center)
CT — Channel Termination (flat rate billing)
IOC — Interoffice connection on LDS network (distance-sensitive)
LATA — Local Access and Transport Area (telco service zone)
LDS — Long-Distance Service (common carrier)
LEC — Local Exchange Company (local telco)
POP — Point of Presence [LDS's office in each LEC: also called SO (Serving Office)]

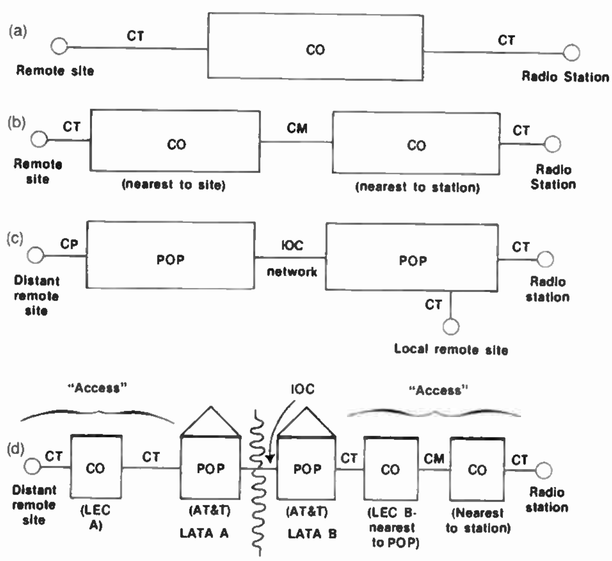


Figure 3.11-8. Billing methodology for dedicated telco circuits. In (a), the two ends of a local circuit are served by the same central office, or "rate center." In (b), two different central offices are involved. In (c) and (d), long distance service is depicted. LEC connections to LDS are referred to as "access." Access shown in LATA B will be more expensive because the radio station is in different rate center of LEC B than the long distance carrier's POP.

AMI. Alternate Mark Inversion. The binary modulation code used by the telephone company for data and digital voice transmission. It uses RZ coding in an alternate bipolar scheme, with logical zeros corresponding to zero volts, and logical ones alternating between +3V and −3V. (The first logical 1 produces a +3V output, the next 1 produces −3V, the next +3V, and so on.) Self synchronization is possible with this approach, but the number of continuous zeros must be limited.

B-channel. In ISDN service, a channel designated for customer data transmission, uninterrupted by any signalling data.

Bellcore. Bell Communications Research. The R&D firm that feeds technology and standards to the RBOCs, and is funded by them.

Carrier. In telco parlance, refers to a multiplexed digital interoffice signal, containing many individual calls or signals in a single cable or fiber.

CCITT. Consultative Committee of International Telephone and Telegraphy. The international standards setting organization for telephone systems, established by the UN.

Codec. Coder/decoder. Any device which includes digital transmission/encoding and reception/decoding circuitry in the same chassis.

CPE. Customer Provided Equipment. Refers to any network interface hardware not provided by telco.

CSU. Channel Service Unit. Terminal hardware for a telco data line, either CPE or telco provided. Also referred to as CSU/DSU (DSU = Data Service Unit) in T1 applications. Interfaces unipolar NRZ computer style datastreams to the RZ bipolar (AMI) telco data format. A switched CSU includes a keypad for call direction and other switch control.

D-channel. In ISDN service, a channel designated for signalling data only.

DDS. Dataphone Digital Service. The first telco data service in the U.S., originated in the mid 1970s by AT&T.

DS0. Digital Service 0. A 64 kb/s data channel.

DS1. Digital Service 1. A 1.544 Mb/s data service usually configured as 24 DS0 channels plus an 8 kb/s sync channel.

DS2. Digital Service 2. Four DS1 channels multiplexed together for transmission. Generally reserved for telco interoffice transmission, and not offered to customers directly.

DS3. Digital Service 3. Twenty eight DS1 channels multiplexed together with additional control data, providing a data rate of 44.736 Mb/s (generally quoted as 45 Mb/s). Used for compressed NTSC and high definition video distribution.

First-Mile. Refers to the signal path between a program's origination site and its entry point to a common carrier's network or a private satellite uplink. Usually a terrestrial RF link or a local telco circuit.

G.722. A CCITT standard for audio data compression. It uses two-subband ADPCM coding to put 7.5 kHz audio into 64 kb/s.

HDSL. High-speed Digital Subscriber Link. A digital service that uses telcos' existing copper (twisted-pair) infrastructure to provide high-speed bidirectional connectivity over limited distance. Adaptive terminal hardware allows higher speed than previous equipment was capable of on the same physical paths. Bidirectional rates of 6 Mb/s or higher are possible. The smart hardware allows quick installation without extensive testing and tweaking of lines. HDSL is replacing T-1 service in some areas.

InterLATA. Refers to telco service or rates between LATAs, or long distance service.

IntraLATA. Refers to telco service or rates within a LATA, or local service.

ISDN. Integrated Services Digital Network. A new telco service designed to eventually replace POTS with flexible digital service. It will be offered in basic rate (2 B + D) service, intended for home use, and primary rate (23 B + D) service for business customers.

J.41. A CCITT standard for digital audio encoding. Using 1411 PCM encoding (14 bits for lower level signals, 11 bits for higher level signals), it places 15 kHz audio on 384 kb/s.

Last mile. Refers to the short haul signal path between a long distance network terminal point (or private satellite downlink) and the customer's receive point. Usually a local telco circuit.

LATA. Local Access and Transport Area. The service area of a local exchange company (LEC).

LDS. Long Distance Service. A carrier of long distance (interLATA) telecommunications, such as AT&T, MCI, Sprint, and others.

LEC. Local Exchange Company. A local telco. Each RBOC contains one or more LECs. Also refers to independent, non Bell local telcos.

Mark. The telco term for high level data pulse, usually corresponding to logical 1. (See Space).

NRZ. Non-return-to-zero. The most basic form of binary modulation coding, in which logical 1s and 0s are directly represented by high and low levels respectively. Because no level transition occurs between continuous strings of like logical values, this form of modulation is not self synchronizing, and requires an external bit clock output for synchronous operation.

Packet switching. A sort of data partyline, in which data is transmitted in addressed bursts or packets, occupying the transmission channel only for the duration of the packet, after which the channel is free for other packets to or from the same or other users. Many users can be interconnected to the same line, but data can be sent discretely to each destination.

PDN. Public Data Network. Telco data services, including both switched and leased lines.

POTS. Plain Old Telephone Service. Refers to the public switched telephone network (PSTN).

PSTN. Public Switched Telephone Network. The standard dial-up phone system.

RBOC (or BOC). (Regional) Bell Operating Company. The seven Baby Bells created when AT&T divested itself of its local telephone operations.

RZ. Return-to-zero. A form of digital modulation coding in which logical ones and zeros are directly represented by high and low levels respectively, but where coding output returns to low level following each high pulse.

Slot (or Time-slot). Generally refers to a DS0 channel within a DS1 signal.

SMDS. Switched Multi Megabit Data Service. A high speed switched data network, operating at DS1 to DS3 rates.

SONET. Synchronous Optical Network. A telco standard for ultrafast data transmission, operating at speeds of 150 Mb/s to 2.4 Gbit/s.

Space. The telco term for low level data pulse, usually corresponding to logical 0 (see Mark).

Switch. Generic name for any telco call routing and connection hardware.

Switched 56. A switched digital service offering 56 kb/s data service on a dial up network.

TA. Terminal Adapter. The term for a CSU in an ISDN system.

Tariff. A schedule of services and their prices that a telco will provide to a given service area, subject to approval by the appropriate regulatory agency.

T-carrier. See Carrier.

T1. The copper network and hardware used to carry DS1 service.

VSAT. Very Small Aperture Terminal. Refers to Ku band satellite earth stations for fixed or portable use with dish diameters on the order of 1.5 m or less.

V.35. An older (but still common in the U.S.) CCITT telco standard for low speed data I/O to a CSU, with a unique multipin connector.

xDSL. Generic terminology for any of the variety of Digital Subscriber Link services offered by telcos, which provide high-speed connectivity across limited distances on the existing copper (twisted pair) infrastructure. (See ADSL, HDSL.)

Section 4: Radio Transmission Facilities

Introduction by Milford Smith, Greater Media, New Brunswick, NJ

Before radio is anything else it is, at its very essence, the transmission of electromagnetic waves from one place to another. Before we can play the hits, talk the talk or spread the news, the ability to manufacture, impart information to and launch these invisible wavelengths is essential and vital. Ironically, as time has passed and the machines associated with the radio frequency portion of radio stations have become infinitely more reliable, the relative level of expertise of the typical radio engineer in the RF domain has arguably diminished. This is in no way reflective of the overall expertise of the individual nor their desire to be knowledgeable in this important area, but is simply a result of the reality of today's staffing levels and the fact that other portions of the transmission chain are usually more demanding of immediate attention. It is for this reason that the material presented here is of such great value.

In the following chapters, experts from virtually every area of radio frequency transmission share their unique expertise with those of us charged with the day to day maintenance, trouble shooting upkeep and regulatory compliance of AM and FM RF plants.

The section begins with George Woodard's skillful discussion of the viability of AM transmission systems in the *AM Transmitters* chapter.

In the *AM Stereo* chapter, Edward J. Anthony discusses the more practical aspects of AM stereo preparation, installation and maintenance to aid the AM broadcaster in achieving a high quality stereo transmission.

Geoff Mendenhall's chapter, *FM Broadcast Transmitters,* deftly describes the advantages of stereo multiplex FM broadcasting. While, John Kean presents the reasons why most broadcasters agree that FM's resurgence and eventual dominance as an audio medium was due in large part to stereophonic transmission in his chapter, *FM Stereo and SCA Systems.*

Jerry Whitaker discusses the importance of the studio-to-transmitter link as digital technology becomes more prevalent in everyday radio operations in the *STL Systems* chapter. Continuing in the realm of digital technology's effect on radio broadcasting, Scott Wright unravels the mystery of the "chicken and the egg" problem that exists between radio broadcasters and RBDS manufacturers in his *Radio Broadcast Data System (RBDS)* chapter.

Harold Hallikanien walks novice and expert alike through the various meters and controls that the Federal Communications Commission requires broadcast transmitters to have in his *Transmission System Control and Monitoring* chapter.

Carl E. Smith's *AM Broadcast Antenna Systems* chapter provides station engineers with an understanding of some of the basic concepts of antenna design. Smith's chapter is followed by Ed Edison's *AM Antenna Coupling and Phasing Systems* chapter, which focuses on the function of such systems and their performance objectives, as well as the characteristics of basic networks, power dividers, transmission lines, sampling systems, detuning systems and transmitter load optimization.

James B. Hatfield's *Computer Simulation of AM Radio Antenna Systems* chapter teaches radio broadcasters how modeling an AM array can save a great deal of trial and error.

In the *Maintenance of AM Broadcast Antenna Systems* chapter, Ed Edison and Gerhard J. Straub team up to discuss the importance of maintenance attention in overcoming the normal effects of age and deterioration (routine maintenance), restoring proper operation after catastrophic failures (such as caused by lightning) and coping with new sources of reradiation that may develop within the environment in which the antenna system operates.

Written by Peter K. Onnigian, the *FM Broadcast Antennas* chapter is for broadcast engineers, technicians and station managers who must make important decisions regarding FM transmitting antennas.

Robert A. Surette details how filters and other components in FM combiners are used in transmitting several frequencies from a single broadband antenna system in the *FM Combining Systems* chapter.

In the *FM Translators and Boosters* chapter, Chip Morgan points out how proper translator and booster design and optimization methods can make the difference between a successful operation and a failure.

Finally, the section concludes with Donald Everist's chapter, *AM and FM Field Strength Measurements,* which explains the field strength measurements procedures for both types of stations.

The material presented is practical and ultimately usable by those of us in the trenches.

4.1 AM TRANSMITTERS

GEORGE W. WOODARD, P.E.
INTERNATIONAL BROADCASTING BUREAU, WASHINGTON, DC

INTRODUCTION

Broadcasting to the general public began with the process known as *amplitude modulation (AM)* near the start of the second decade of the twentieth century. Then, as now, the system of modulation chosen for transmission was heavily dependent on practical and economic aspects of receiver technology. The evidence is clear that amplitude detection was the only known practical method of signal demodulation when the ideas of radio communication began to be formulated in the late 1800s and early 1900s. It is apparent from some of the earliest technical writings on radio communications that a practical understanding of angular modulation (frequency and phase modulation) did not exist until the mid-1920s.[1] These modes of transmission and the necessary receiver technology were not to be proven practical until long after AM had become the standard of the radio and broadcasting technological art. AM in broadcasting is today more accurately referred to as DSB-FC-AM, an acronym for double sideband full carrier amplitude modulation, however, the abbreviation AM has been generally accepted to describe the mode of transmission currently used throughout the world in the standard broadcast AM band of 535 to 1705 kHz. In North America (ITU Region 2), the channel separation on the AM band is 10 kHz. In much of the rest of the world (ITU Regions 1 and 3), the channel separations are 9 kHz. In most of the rest of the world outside North and South America (ITU Regions 1 and 3), the AM band is more commonly known as the medium wave or MW band. The long-wave or LW band of 150 to 280 kHz, still popular in parts of ITU regions 1 and 3, also uses AM.

From almost the very beginning of radio communication, it was known that AM was not a very efficient mode of transmission, either in spectrum usage or in transmission of intelligence. The pioneers in AM broadcast transmitter and receiver engineering technology had a deep sense of responsibility toward developing this marvelous medium to enhance the lives of the general public with informative, educational, and entertaining programs. Some of the early communication technology pioneers in the U.S. and abroad had great vision for both the technical and programming aspects of public broadcasting.[2]

In the 1930s, when the telephone industry was planning a major switch from DSB-FC-AM in favor of single sideband suppressed carrier (SSB-SC) for both its long distance wire and wireless services because of its higher transmission efficiency and channel capacity,[3] the broadcast industry was committed to the continuance of conventional AM because of the economic need for public receiver compatibility. Therefore, the broadcast engineers' attention naturally turned to improvement of transmission standards such as fidelity, efficiency of transmission, reliability of transmission, and co-channel and adjacent channel interference, etc. Of these, efficiency and reliability of transmission have seen and continue to see the greatest improvement in the views of many transmitter designers and users. Transmission fidelity, always important, reached a plateau in the late 1940s of approximately 2–3% total harmonic distortion (THD), 10–15 kHz audio bandwidth, and −55 dB noise, which had not been significantly improved upon until the early to mid-1980s when totally solid-state transmitters began to appear on the AM transmitter market, yielding nominally 1% or less THD and −65 dB or better noise. In addition, pretransmitter audio program processing equipment and philosophy, which influence perceived reception fidelity, continue to improve and change through use of modern linear and digital circuit design techniques.

In the early 1980s, stereo transmission on the AM band became a reality. Some AM broadcasters saw AM stereo as a necessary means to compete with FM stereo. AM stereo transmission was approved on a general basis by the FCC in 1981, but unlike FM stereo or color television, the FCC did not adopt a transmission standard for this potentially revolutionary development in the standard broadcast band. Instead, the FCC allowed any commercially developed AM stereo transmission system, which would pass FCC type approval requirements regarding mono compatibility and interference, to compete for the broadcasters' and consumers' favor in order to establish a *defacto* standard for AM stereo transmission. This policy ultimately failed because receiver manufacturers could not justify the cost and complexity of receivers capable of receiving multiple AM stereo formats, and very few AM stereo receivers made it into the marketplace. Without many receivers, broadcasters had little incentive to begin or continue AM stereo transmissions. The FCC eventually recognized the failure of its policy and adopted a single AM stereo standard. However, as of this writing there are only a limited number of broadcasters and receiver manufacturers using this standard in the United States.

AMPLITUDE MODULATION THEORY

Amplitude modulation of a radio signal occurs in at least two basic forms, coded (digital) and linear (analog). The first AM processes for long distance communication involved on/off keying of a radio carrier wave. The pattern or *code* of the on/off keying process determined the content of the information being transmitted. In current state-of-the-art, linear or quasi-linear undulations are used to transmit the analog information present in speech and music. The radio carrier wave signal onto which the analog amplitude variations are to be impressed is expressed as:

$$e(t) = A*Ec*Cos(wct) \quad [1]$$

Where:

$e(t)$ = instantaneous amplitude of carrier wave as a function of time (t)
A = a factor of AM of the carrier wave
wc = angular frequency of carrier wave (radians/second)
Ec = peak amplitude of carrier wave

If A is a constant, the peak amplitude of the carrier wave is therefore constant and no modulation exists. Periodic modulation of the carrier wave exists if the magnitude of A is caused to vary with respect to time as, for instance, a sinusoidal wave.

$$A = 1 + (Em/Ec)*Cos(wmt) \quad [2]$$

Where: Em/Ec is the ratio of modulation amplitude to carrier amplitude
leading to:

$$e(t) = Ec*(1 + (Em/Ec)*Cos(wmt))*Cos(wct) \quad [3]$$

this is the well known basic equation for periodic (sinusoidal) AM and when all multiplications and a simple trigonometric identity are performed the result is:

$$\begin{aligned} e(t) = {} & Ec*Cos(wct) \\ & + (M/2)*Cos(wct + wmt) \\ & + (M/2)*Cos(wct - wmt) \quad [4] \end{aligned}$$

Where: M = the AM factor Em/Ec

Equation [4] can be represented graphically in at least two practical ways: in the time domain representation as shown in Figure 4.1-1, and in the frequency domain as shown in Figure 4.1-2. The graphical representations shown are for a single tone modulation index (M) of 0.7 (the peak modulating voltage is 70% of the peak carrier wave voltage (Em/Ec = 0.7)). Figure 4.1-2 shows the occupied bandwidth of an AM signal with single tone modulation. From this figure and its defining equation [4], it is clear that the bandwidth of an AM signal is equal to twice the highest modulating frequency when no system distortion is present. High quality music reproductions include frequency components as high as 15 kHz or higher and therefore the required theoretical bandwidth of a DSB-FC-AM signal capable of high quality music reproduction is at least 30 kHz. Harmonic and intermodulation system distortion have the effect of widening the effective occupied bandwidth of the system. However, most modern transmitters have sufficiently low distortion characteristics that bandwidth stretching due to system non-linear distortions is not normally a significant problem. The occupied bandwidth characteristics of an AM broadcast transmitter are discussed in more detail in the section on *factory tests* and *audio processing* and *pre-emphasis*.

Digital Radio Broadcasting

At this writing, research is well underway on several techniques to transmit digital audio broadcasting on the LW, MW, and SW-AM bands, as well as on VHF-FM bands. The technology for both compatible and discrete AM and FM digital broadcasting is still in its infancy and largely proprietary to the many competing proponents.

It is clear that digital radio broadcasting will soon be a reality with great promise for increased quality and reliability of signal delivery to the listening public. Because of the tremendous global economic investment in transmitting equipment, it is likely that any successful digital radio broadcasting system will, at least initially, use existing transmitting equipment in some relatively easily modified form. All proposed digital radio broadcast systems hold the promise of equal or greater current performance and coverage with considerably less than current transmitting power output levels.

RADIO FREQUENCY POWER AMPLIFIERS

The most important system component which is common to all AM systems is the high power final RF power amplifier. High power amplifiers that produce 0.25 to 50 kW of carrier power are common for AM broadcast transmitters in North America. Carrier power levels up to 1 MW and higher are not uncommon in other parts of the world for medium wave broadcasting. Transmitters delivering these high power levels should be designed for highest operating efficiency within the circuit constraints of the particular manufacturer and/or modulation system used. The most common amplifier used to meet the demands of high output power and high efficiency in modern transmitter designs is the class "C", "C/D", or "D" amplifier, either vacuum tube for large output powers from a single amplifier circuit or solid-state for large output powers derived from several combined lower power amplifier circuits. Class "B" RF power amplifiers, very popular in early broadcast transmitters, are still used occasionally at lower power levels and for driver stages of final class "C" or "D" stages. Solid-state class "D" amplifiers up to 5 kW are becoming the standard as driver stages for final vacuum tube amplifiers and for the final power amplifier modules in modern all solid-state AM transmitters. As stated in the introduction, the most common major concerns for both manufacturers and users of modern AM broadcast transmitters is operating relia-

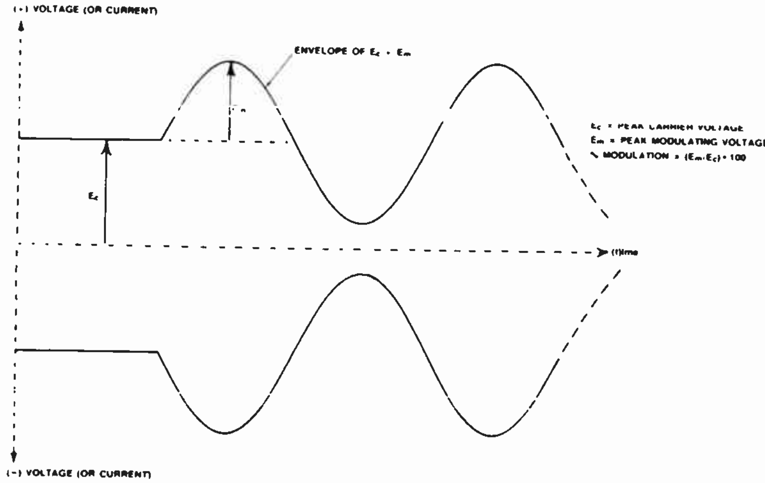


Figure 4.1-1. Time domain representation of a carrier wave signal amplitude.

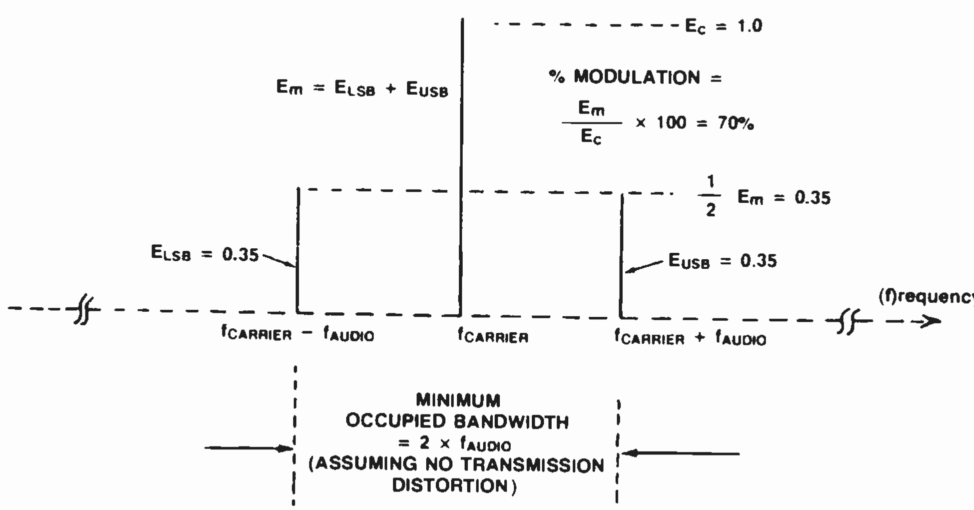


Figure 4.1-2. Frequency domain representation of an amplitude modulated signal showing the carrier wave signal and two resultant modulation sidebands at 70% modulation.

bility and efficiency, hence operating cost. To achieve high overall operating efficiency, the stages which consume the most power, the modulator and/or the final RF power amplifier stages must be designed for highest possible operating efficiency.

The basic tuned anode vacuum tube amplifier is described in graphical form in Figure 4.1-3. The vacuum tube shown can be either a triode, tetrode, or pentode, and can also be representative of transistor stages. Tetrode final amplifiers are most common in modern high power transmitter vacuum tube designs. The RF excitation voltage is supplied to the grid of the power amplifier tube and the ratio of dc grid bias voltage-to-peak RF excitation voltage, shown sinusoidal in Figure 4.1-3, determines the conduction angle (θc) of anode current, given as:

$$\theta c = 2*\text{Arc Cos}(Ecc/(Eg - Ecc)) \qquad [5]$$

where the exciting grid signal (Eg) is sinusoidal as shown in Figure 4.1-3(a).

The shape of the anode or transistor collector current pulse is determined by the device transfer characteristics and input waveshape. The pulse of current thus generated, Figure 4.1-3(c), is supplied by the dc power supply, E_{BB}, and passed through the resonant tank circuit shown in Figure 4.1-3(d). The resonant tank circuit is assumed to have sufficient operating "Q" to force anode or collector voltage (ep) to be essentially sinusoidal and of the same periodic frequency as the RF excitation voltage and resultant anode or collector current pulse. The instantaneous anode dissipation, shown in Figure 4.1-3(e), is the product of instantaneous tube anode voltage drop and anode current. The tube transfer characteristic is a variable dependent upon many tube factors as well as maximum drive signal (Eg). The exact shape and magnitude of the current waveform is

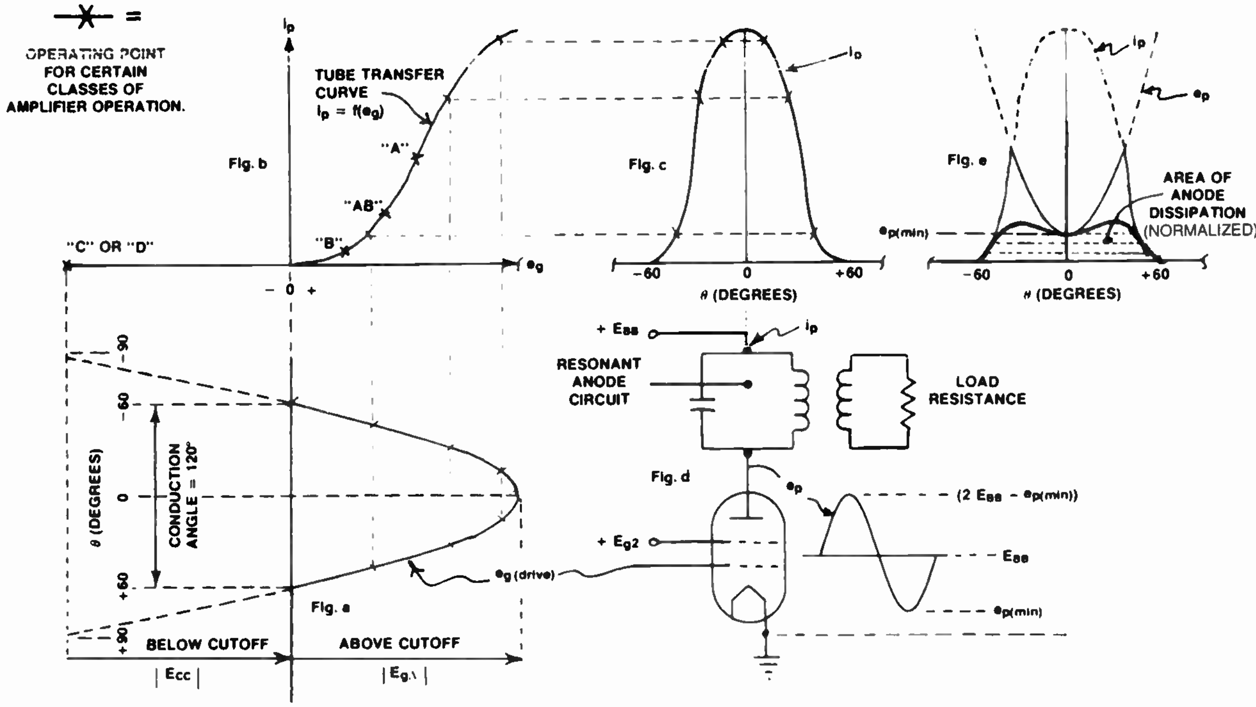


Figure 4.1-3. (a), (b), (c), (d) Classical vacuum tube class "C" amplifier with sinusoidal grid drive, 120° anode current conduction and resonant anode load.

normally obtained from a load line plot on constant current characteristic tube curves supplied by the tube manufacturer. The resonant anode load impedance is chosen and adjusted to allow ep (min) to be as low as possible without causing excessive screen grid (in the tetrode case) or control grid dissipation. Some manufacturers increase anode efficiency beyond the limits for typical Class "C" amplifiers by using a circuit employing a third harmonic resonator between the output anode connection and the fundamental resonant circuit. This has the effect of squaring up the anode voltage waveform (ep) thus causing the integral of the ep × ip product, or anode dissipation, to be smaller; resulting in lower anode dissipation for a given RF power output. An amplifier employing the third harmonic anode trap is commonly referred to as class "C-D;" suggesting an efficiency rating somewhere between conventional class "C" operation (nominal 120° conduction angle) and true class "D" operation with rectangular anode or collector voltage waveforms. Anode and collector efficiencies can be increased typically to values of 90% for transmitters up to approximately 10 kW carrier power and approximately 85% for transmitters higher than 10 kW carrier power by using the third harmonic trap technique. Table 4.1-1 shows a comparison of anode efficiency for six classes of high power tuned RF amplifiers under certain defined conditions.

Table 4.1-1

Comparison of Tuned RF Amplifier Efficiencies

Amplifier Class	Conduction Angle (degrees)	Anode Efficiency (%)	Defined Conditions of Operation
A	360	30	$E_{b(min)} = 0.10 \cdot E_{BB}$
A-B	240	60	$E_{b(min)} = 0.10 \cdot E_{BB}$
B	200	67	$E_{b(min)} = 0.10 \cdot E_{BB}$
C	120	84	$E_{b(min)} = 0.05 \cdot E_{BB}$
C-D	120	90	$E_{b(min)} = 0.05 \cdot E_{BB}$ (Odd Harmonic Resonators)
D	120	95	$E_{b(min)} = 0.05 \cdot E_{BB}$ (Square-wave voltage shape)

High Level Anode Modulation

The first practical method of generating the AM signal was Heising constant current, a method of applying audio modulation to the anode supply voltage of a class "C" RF amplifier.[4,5] This general type of modulation has since been known as high level anode (or plate, or collector) modulation (see Figure 4.1-4). The Heising modulator was used at least as early as 1920 and was usually used to modulate a low power RF amplifier or master oscillator stage which was followed by several linear amplifier stages until the desired power level was attained. In some cases the Heising modulator was used to modulate the final RF amplifier stage of lower power transmitters. The Heising shunt modulator operated in the class "A" mode and therefore was low in operating efficiency. The early linear amplifiers were tuned class "B" amplifiers operating with carrier level anode efficiencies of 30% maximum. The Heising and similar systems of audio amplification were also used to modulate the grid bias level of RF amplifier stages in order to obtain the

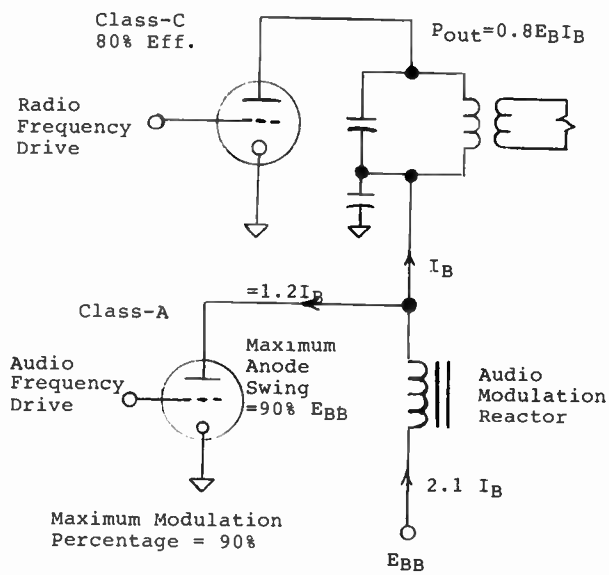


Figure 4.1-4. Heising Constant Current Modulator.

AM signal to be used for further linear amplification. Heising constant current anode modulation was also very popular in military and aviation radio sets used through the end of World War II.

Chireix Outphasing Modulation

Outphasing modulation was originally described in the literature by its inventor Henry Chireix in 1935.[6] It is a unique and ingenious method of obtaining the AM signal by use of counter phase modulation and vector addition of two separate radio frequency signals. It was marketed for many years by RCA under the trade name Ampliphase and many of those transmitters and some of European manufacture are on the air still today at power levels of 100 kW and higher. This system of modulation is described graphically in Figure 4.1-5. Two RF signals are derived from a common excitation source and then split into two separate channels. Each channel is shifted in phase, one positive and the other negative. The two channels are then each phase modulated by the modulating signal, again, in opposing polarity. The two channels are amplified and then recombined in a vector additive network which has the effect of producing the desired AM. The main advantage, as with all systems previously discussed, is in operating efficiency. The two independent channels contain only phase modulated RF signals and therefore each can be amplified to the desired power levels in high efficiency Class "C" or "D" amplifiers. The actual modulation process takes place both at low level, in the phase modulators, and at high level, in a passive output network combiner.

There are two major disadvantages of this system of modulation. First, the efficiency of the output power amplifiers is not quite as high as the previous description would imply, due to the fact that at all instantaneous levels of modulation, except one, the anode circuits must work into a reactive load. Secondly, output carrier power setting is sensitive to tuning of any stage anywhere down the chain; operators of early versions of this type of equipment soon learned of the potentially disastrous consequences of trimming a tuning control of some lower power stage. Later designs used broadband amplifying stages for the lower levels to circumvent the problem of tuning sensitivity of the lower power stages.

Another major disadvantage of the Chireix system in older transmitters was the generation of incidental phase modulation (IPM) in excessive amounts when the two low level phase modulators were not well balanced in opposing PM characteristics. It was common for such an unbalance to exist producing as much

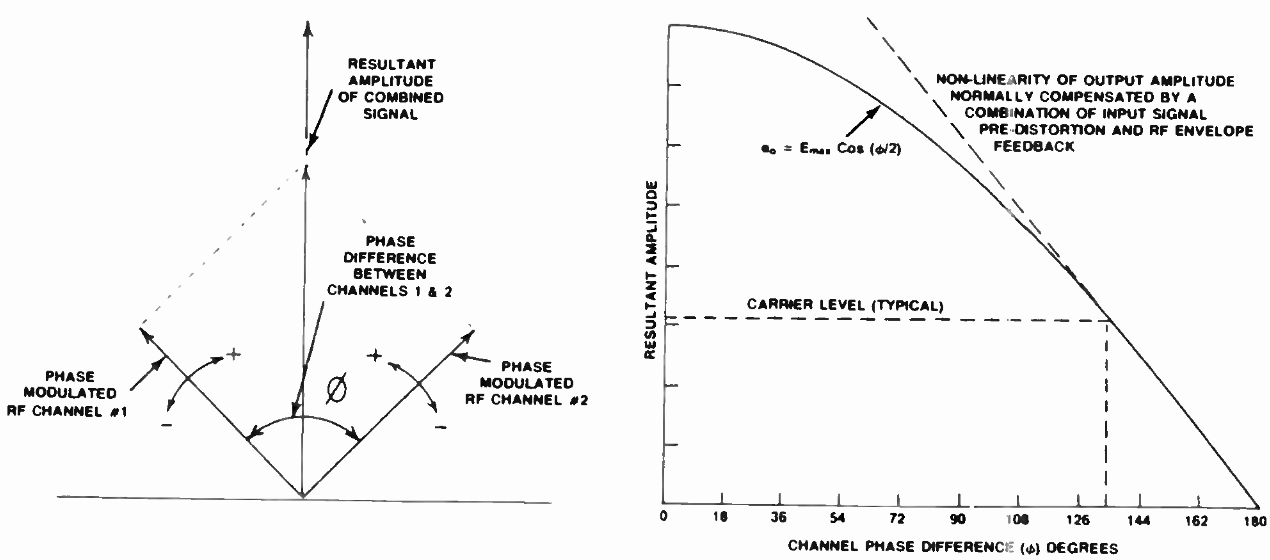


Figure 4.1-5. Principle of Chireix outphasing modulation.

as 12 to 18° of peak IPM at either 100% modulation crest or trough. Modern solid-state circuitry for the phase modulators using digital phase modulation techniques would completely remove this disadvantage were the system still being commercially produced. The outphasing system of AM ceased to be produced commercially when RCA discontinued its AM transmitter manufacturing in the mid-1970s.

Class "B" High Level Anode Modulation

Historically the most popular method of applying the audio modulating voltage to the anode circuit of a class "C" RF power amplifier was by a high power push-pull class "B" audio amplifier. This type of modulation was first used to improve the operating efficiency and to increase the output power of AM broadcast transmitters. Class "B" push-pull audio amplification was first used to improve distortion and output power of telephone transmission amplifiers. The invention was soon recognized by broadcast engineers and applied to high level anode modulation. With the final RF power amplifier operating at approximately 80% anode efficiency and the class "B" audio modulator total static currents approximately one-tenth that of an equivalent Heising modulator, total anode efficiencies at carrier level rose to approximately 72% compared to 37% for the Heising system and 30% for conventional linear amplification. A simplified drawing of a typical high level class "B" anode modulation system is shown in Figure 4.1-6. The vacuum tubes shown in Figure 4.1-6 may be either triodes, tetrodes or pentodes. The output circuit of the class-B modulator shows the output modulation transformer (MT), an audio coupling capacitor (C), and a dc shunt feed inductor (L).

This arrangement was used in all high level class-B high power broadcast transmitters until about 1960 because of a transformer design constraint that would not economically allow unbalanced direct currents to magnetize the transformer core material without poor low frequency distortion. Advanced technology core materials and careful magnetic transformer design allowed elimination of the coupling capacitor and the dc feed shunt inductor, first in some 100 kW European transmitter designs in the early 1960's, and in an American shortwave transmitter design in the later 1960's. Many of the more advanced modern AM broadcast transmitter designs still using high level class-B anode modulation have eliminated the extra C and L components from the modulator output circuitry, and the dc current to the modulated RF amplifier anode flows directly through the secondary of the output modulation transformer. Elimination of these C and L components is necessary for optimum operation of modern AM stations. With the extra C and L components, the modulator output is effectively a three-pole high pass filter which causes low frequency transient distortion to be generated when driven with the complex waveforms that are produced by many modern and popular program audio processors. Eliminating the C and second L component, causes the output modulator circuit to become a single-pole high pass filter, greatly reducing low frequency transient distortion.

Another problem with transformer coupled high level class-B anode modulation is with high frequency audio transient distortion. Stray internal winding capacitances and leakage inductances form multi-pole low pass filtering at the high frequency end of the audio spectrum. This equivalent multi-pole low pass

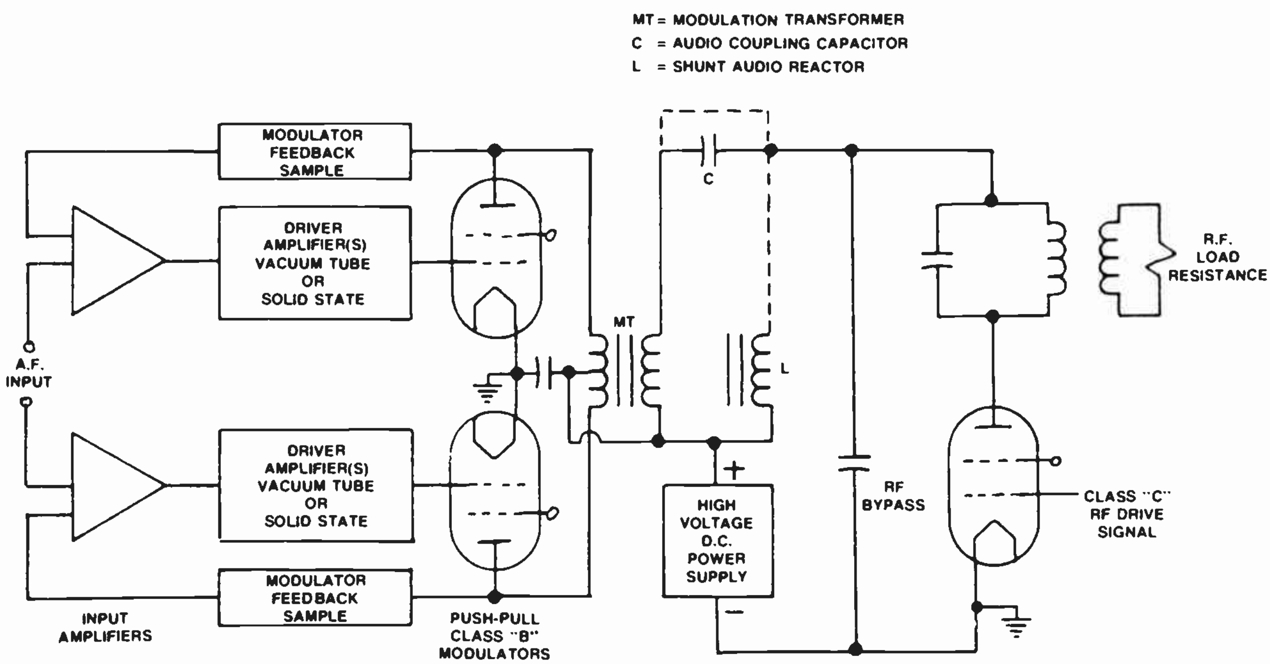


Figure 4.1-6. High Level anode modulation employing transformer coupled push-pull class-B modulators.

filter generates transient overshoot distortion when driven by the same type of processed complex program waveforms previously mentioned. Transient overshoot up to 12% is typical for square wave modulator input signals and results in a required modulation level reduction of the same 12% in order to maintain peak modulation levels within FCC allowed limits. This high frequency transient overshoot distortion can be effectively minimized by filtering the audio input to the transmitter with linear phase filters, resulting in somewhat lower high frequency audio response, and/or by careful control of the modulation transformer equivalent circuit yielding more linear audio phase characteristics for the entire modulator circuitry. Balanced modulator negative feedback is used to reduce modulator non-linear distortion and noise. Negative feedback, however, usually worsens high audio frequency transient distortion characteristics.

Doherty High Efficiency Linear Amplifier

The Doherty High Efficiency Linear Amplifier was first described in the technical literature in 1936 by its inventor, W.H. Doherty.[7] So contrary were the terms *linear* and *high efficiency*, in the context of amplitude modulated waves, that many engineers in broadcasting refused to accept the concept as workable, similar to the reaction Armstrong received when he proposed that frequency modulation was a practical mode of radio transmission. Nevertheless, the Doherty High Efficiency Linear Amplification System was soon proven to work by 1938 and has been used at power levels up to 1,000 kW carrier power in both the original and in the patented Weldon modified form in many installations throughout the world on the medium wave broadcast and international shortwave broadcast bands, as well as the long wave broadcast band in Europe.[8,9] Its implementation was the result of true inventive genius, using one or more known basic scientific principles to create a totally new and necessary product. The Doherty Linear Amplifier is described graphically in Figure 4.1-7. As with conventional linear amplifiers, the AM signal is generated at low levels and applied to the input of the final amplifier stage. The Doherty System employs two output amplifier stages, one defined as the carrier amplifier and the second as the peak amplifier.

The outputs of the two stages are combined in phase at the anode of the peak amplifier tube. At carrier level the carrier tube is operated as a nearly saturated Class "B" amplifier and thus delivers almost all of the carrier power at Class "B" efficiencies (approximately 70% anode efficiency). The peak tube at carrier condition is biased and driven just above cutoff and therefore supplies a small amount of carrier power (approx. 2 to 6%). The anodes of the two tubes are connected together through a 90° impedance inverting RF network. As the modulated signal increases in the positive direction to both peak and carrier tubes, the current supplied to the output load by the peak tube increases. The saturated voltage drop at the anode of the carrier tube remains constant over the entire positive modulation half-cycle, thus causing the current at the output of the interanode 90° network also to be constant during the same positive modulation half-cycle. The rising current from the peak tube anode has the effect of raising the impedance being presented to the interanode network. Since the current from the network is constant, the net effect is an increase in output power from the carrier tube ($I^2 \times R_L$ increases because R_L increases). At the 100% positive modulation crest, both tubes are producing exactly twice carrier power to the load, satisfying the requirement that peak envelope power (PEP) equals $4 \times P_{carrier}$. During the negative half-cycle of modulation, the peak tube is cut-off and the carrier tube behaves as a normal linear amplifier, allowing the envelope power output to drop linearly to zero at the 100% negative modulation trough. The anode efficiency of the Doherty High Efficiency Linear Amplifier at carrier level is more than twice the efficiency of conventional AM class "B" linear amplifiers.

The Doherty Linear Amplifier also has two other important advantages for high power broadcast transmitters. First, and most important, the peak anode voltage at either tube is only about one-half that required for an equivalent carrier power high level PWM or class "B" anode modulated transmitter thus allowing reliability and usable tube life to increase significantly. Secondly, no large modulation transformer or special filtering components are used in the final amplifier stages to cause transient overshoot distortion as previously discussed for Class "B" anode modulation or pulse width anode modulation. The main problems of the Doherty Linear Amplifier are non-linear distortion and an increase in the complexity of tuning. The major sources of non-linear distortion are the non-linearity of the carrier tube at or near the 100% negative modulation crest and the non-linearity of the peak tube at or near carrier level when it is just beginning to conduct. Both sources of distortion are effectively reduced by use of moderate amounts of overall envelope feedback. The tuning complexity problem is usually overcome by experienced and trained operators and simplified tuning procedures aided by built-in test equipment used in modern designs since approximately 1948.

Pulse Width High Level Anode Modulation

Pulse width modulation (PWM) of the dc anode voltage of a class-C RF amplifier was first used in commercial high power broadcast transmitters in Europe in the early 1960s. It was the first commercially successful attempt to significantly improve upon the efficiency of the popular high level class-B modulation system by applying and improving basic PWM concepts that were described by Heising decades earlier.[10] Since this first success, PWM has become a preferred method of high level anode modulation and is employed in several broadcast transmitter designs by several manufacturers worldwide. The basic PWM system and two ingenious improvements to the basic system are shown in Figure 4.1-8. The circuit in Figure 4.1-8(a) graphically describes the basic principle of

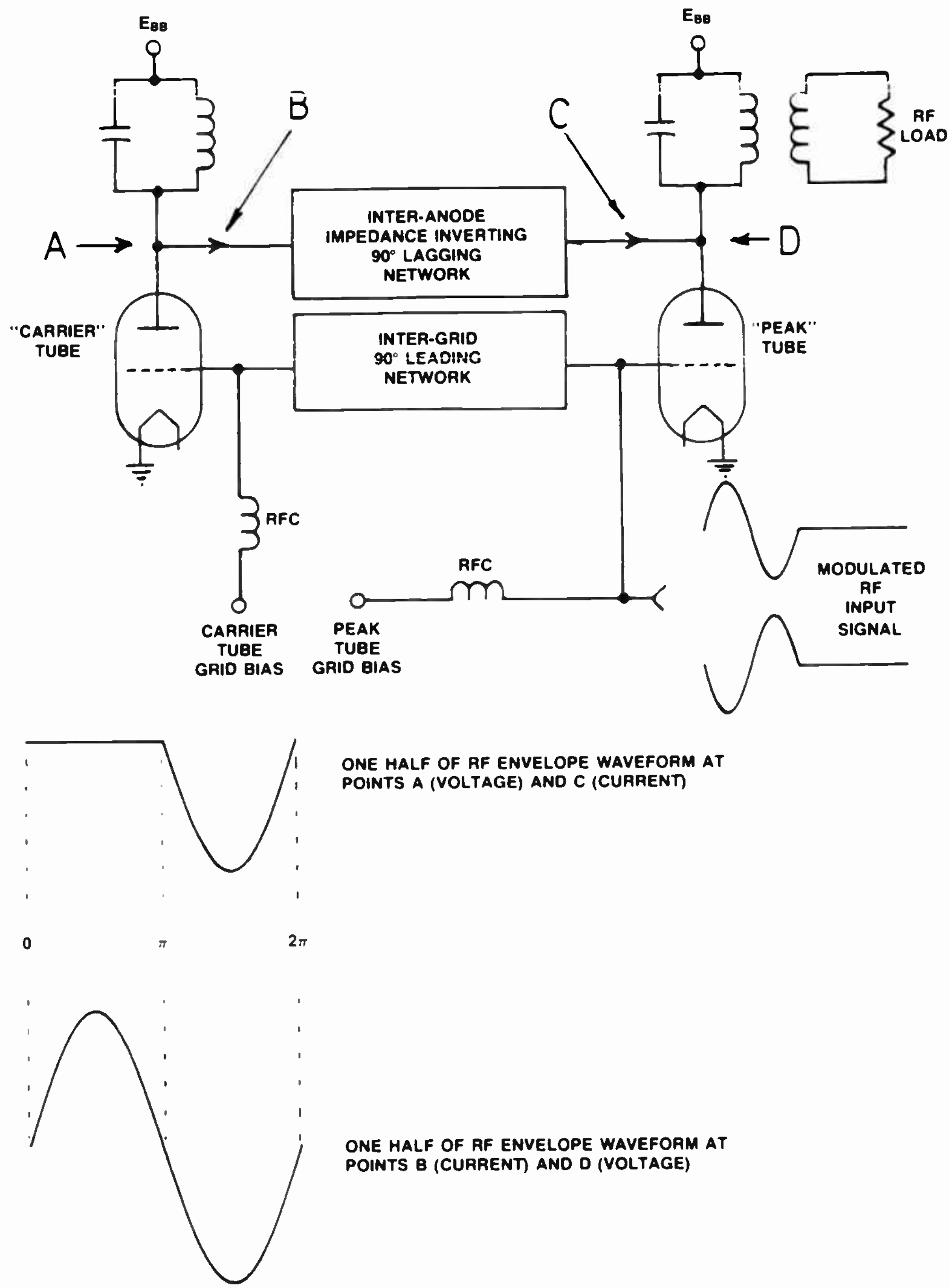


Figure 4.1-7. Doherty High Efficiency Linear Amplifier for modulated waves.

PWM. An inherent practical deficiency in the basic concept, Figure 4.1-8(a), is caused by the relatively high shunt circuit capacitance of the modulator tube filament transformer plus stray capacitances. Though special low capacitance isolation transformers can be used to supply modulator filament and auxiliary power to minimize shunt capacitance, typical realizable values of capacitance can cause excessive switching losses and audio distortion in lower power transmitters.

Besides causing switching losses, the high stray capacitance to ground is also a cause of major modulator distortion at high negative modulation indices.[11] The circuit shown in Figure 4.1-8(b) is one ingenious way to overcome the stray modulator capacitance problem.

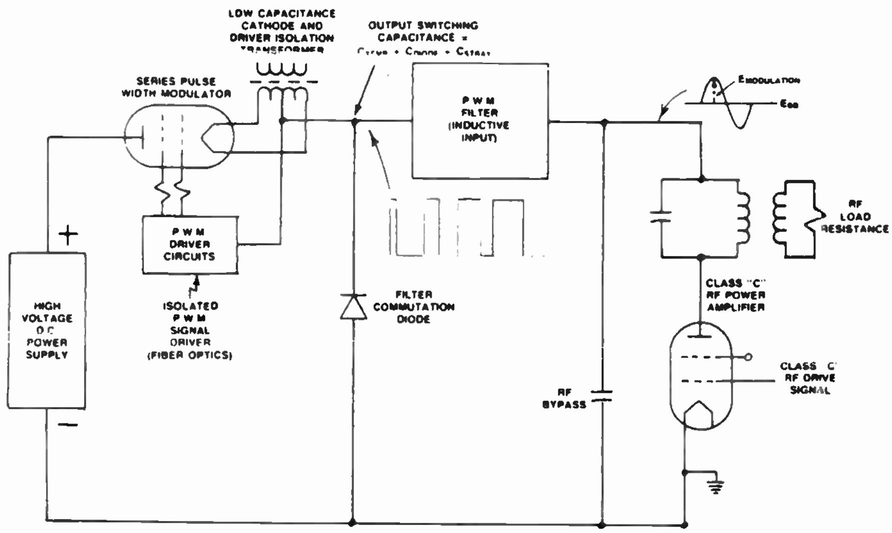


Figure 4.1-8(a). Basic classical high level anode pulse width modulation of a vacuum tube class "C" amplifier.

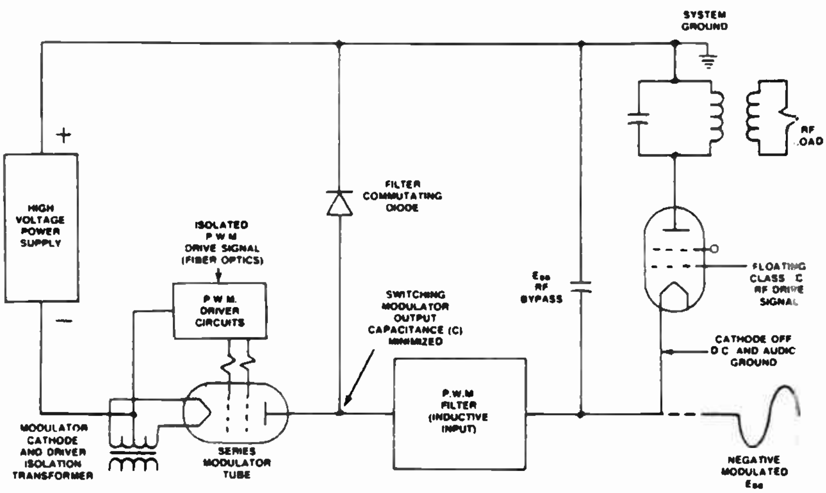


Figure 4.1-8(b). Collins Radio/Continental Electronics patented modification to basic high level PWM System to minimize modulator losses by minimizing switching modulator output capacitance.

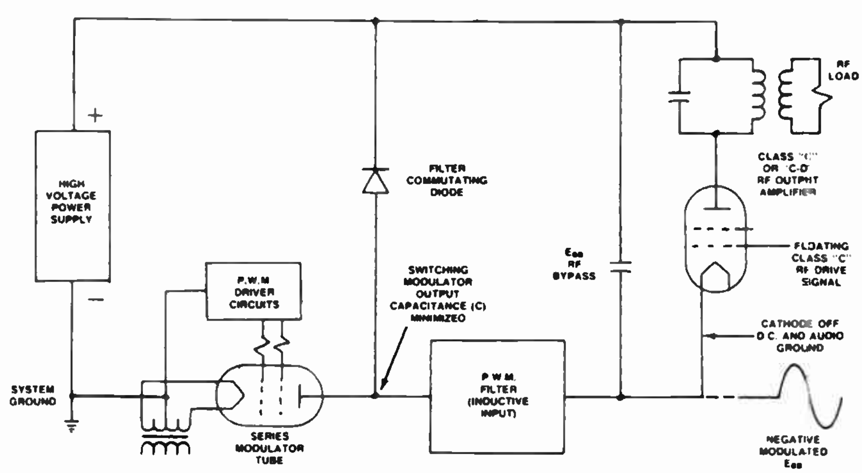


Figure 4.1-8(c). Harris corp. patented modification to basic high level PWM System to minimize modulator losses by minimizing switching modulator output capacitance.

It is identical to the circuit in Figure 4.1-8(a), in principle, except that the high voltage pulse modulated wave is at a point in the circuit where the shunt capacitances to ground are inherently minimized. The circuit in Figure 4.1-8(c) is basically the same as in 4.1-8(b) except the system ground has ingeniously been placed at another point in the circuit. PW Modulator anode efficiencies approach 92% in some higher power transmitter designs yielding a combined modulator and carrier tube anode efficiency of approximately 74% at all levels of modulation. An added efficiency advantage over high level class-B anode modulation is that a PWM transmitter may have only two high power vacuum tubes, one modulator and one final RF amplifier, thus eliminating the filament heating power of one large vacuum tube required in push-pull modulator designs. A significant disadvantage of PWM as described in Figure 4.1-8(b) and 4.1-8(c) is that the cathode and grid circuits of the modulated RF amplifier are operated at high voltage levels off ground, adding complications to the circuitry in these areas that are avoided with the classical PWM circuit of Figure 4.1-8(a) and conventional high level class-B anode modulation. Another major disadvantage of PWM in any form is that of transient distortion caused by the phase non-linearity of the multi-pole PWM filter, similar to that previously discussed for high level class-B anode modulation. The switching frequency is typically 70 kHz for most transmitters manufactured in North America. This frequency is chosen to ease compliance with FCC regulations which requires all spurious radiation more than 75 kHz removed from the carrier frequency to be 80 dB or more below the carrier level. In order to meet the spurious output requirements of paragraph 73.44 of the FCC regulations, a very steep cutoff low pass filter is required at the output of the PW modulator.

High Efficiency Screen/Impedance Modulation

In 1938, Terman and Woodyard described a modification to the basic Doherty High Efficiency Linear Amplification System previously described.[12] In the new system the grid bias level of two tubes operating class "C" is varied at the audio modulation rate thus creating a higher efficiency system of AM rather than amplification while still using the impedance inverting properties of the interanode network described by Doherty. The Terman-Woodyard system of modulation, however, was not extensively used in commercially successful high power transmitter designs.

High efficiency screen/impedance modulation is similar to the Terman-Woodyard modulation system except that the audio modulating signal is applied to the screen grids of two tetrode vacuum tubes operating as class "C" carrier and peak amplifiers. Invented by J.B. Sainton in 1965, the screen/impedance modulation system exhibits significant improvement to the Terman-Woodyard scheme due to the fact that RF excitation voltages and audio modulating voltages are isolated from each other thereby eliminating a troublesome source of tuning versus modulation interaction.[13]

The screen/impedance modulation system is shown in Figure 4.1-9. The peak and carrier tubes are biased and driven in quadrature as Class "C" amplifiers from the continuous wave RF drive source. At carrier level, the screen voltage of the carrier tube is adjusted so that the carrier tube is near anode saturation and delivering approximately 96% of the carrier power; while the screen voltage of the peak tube is adjusted so that the peak tube is just into conduction and supplying the remaining approximate 4% of carrier power. The combined anode efficiency at carrier level is better than 77% as shown in equation [7].

$$n_{at} = 1/(pc/n_{ac}) + (pp/n_{ap})$$
$$n_{at} = 1/(0.96/0.8) + (0.04/0.40) = 0.77 \qquad [7]$$

Where:

- pc = percent carrier power supplied by carrier tube (as a decimal)
- pp = percent carrier power supplied by peak tube (as a decimal)
- n_{ac} = carrier tube anode efficiency at carrier level (as a decimal)
- n_{ap} = peak tube anode efficiency at carrier level (as a decimal)
- n_{at} = total anode efficiency at carrier level (as a decimal)

The modulation of the RF carrier wave occurs when the screen voltage of the peak tube begins to rise during the positive modulation half-cycle, thus causing the peak tube to supply more RF current to the output load. This increase of current into the output network causes the resistance seen by the interanode network to increase and, due to the impedance inverting characteristic of the 90° interanode network, causes a proportional decrease in the load impedance presented to the carrier tube anode. The carrier tube resonant anode voltage drop is fully saturated over the entire positive modulation half-cycle and is therefore effectively a constant voltage source. The power output of the carrier tube thus increases during the positive modulation half-cycle due to the modulated decreasing impedance at its anode until both peak and carrier tubes deliver twice carrier power at the 100% positive modulation crest. During the negative modulation half-cycle, the peak tube is held out of conduction while the carrier tube output voltage decreases linearly to zero output at the 100% negative modulation trough. Simple alterations to the interanode 90° network characteristic impedance gives the screen/impedance modulation system the capability of full FCC allowed positive modulation up to 125% peak without added distortion. The advantages of screen/impedance modulation are the same as mentioned for the Doherty linear amplifier except that screen/impedance modulation has higher efficiency at all depths of modulation and, is less critical to misadjustment of RF amplifier tuning. Screen/impedance modulation has been successfully used in medium wave transmitters throughout the world at the 2 MW carrier power level and up to 250 kW in

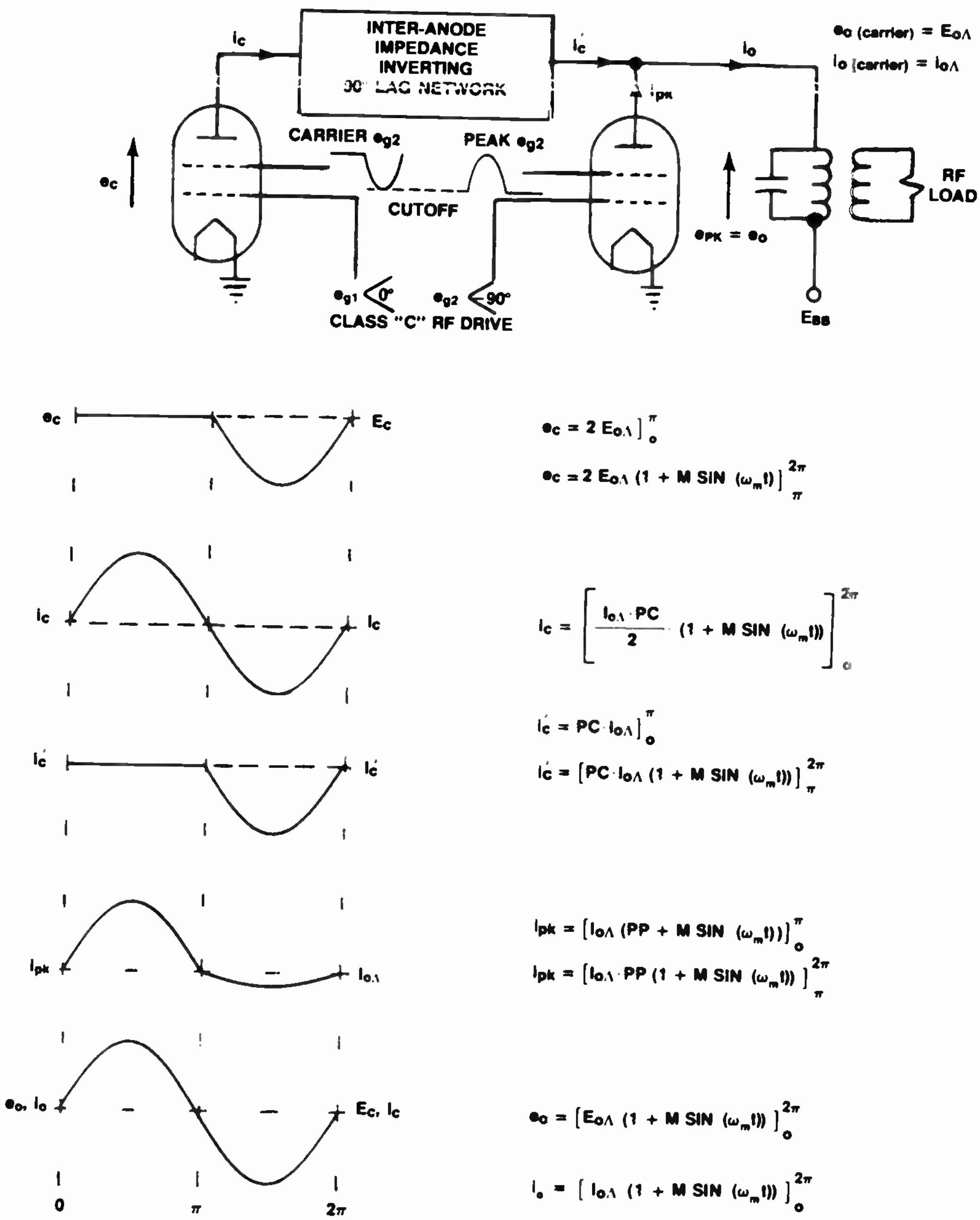


Figure 4.1-9. Principles of Sainton high efficiency screen/impedance modulation.

automatically tuned transmitters used for international shortwave broadcasting.

SOLID-STATE BROADCAST TRANSMITTERS

Developments in the application of solid-state technology to high power medium and shortwave AM broadcast transmitters have been prolific since about 1984 and continue to be a source of great cause for enthusiasm in the high power transmitter technological art. Concerns about the reliability of solid-state circuitry in high power environments has rapidly vanished as the new all solid-state designs and products have proved themselves worthy in the harsh environment of high power transient disturbances caused by gas arcs in associated vacuum tube circuitry and lightning strikes on antennas, power lines, buildings, etc. Current products available on the commercial transmitter market offer solid-state modulators for high level anode modulation of a vacuum tube final RF amplifier up to 600 kW on medium wave and shortwave transmitters. Other products are available utilizing fully solid-state designs up to 1,200 kW for medium wave transmitters. The vacuum tube power amplifier will still have a future for many years as a practical and economical source of radio frequency power generation for shortwave and some very high power lower frequency applications. However, the new all solid-state designs in the medium wave AM bands are achieving great commercial success because of their higher operating efficiencies, reliability, size, weight and superior modulation characteristics over competitive vacuum tube designs. Likewise, products and designs employing solid-state technology in the modulator circuitry for high level anode modulation of power grid vacuum tubes are also enjoying great commercial success worldwide for both medium wave and especially shortwave AM broadcast transmitters.

Solid-State Equipment and Circuitry

Solid-state designs of AM broadcast transmitters currently available on the commercial market fall basically into three categories: solid-state modulator circuitry supplying the audio modulating power for high-level anode modulation of a power grid vacuum tube RF power amplifier, solid-state high level collector modulation of solid-state RF power amplifiers and at least one unique product and design is commercially available in which the AM process is achieved by pseudo digital pulse-code modulation taking place directly in the RF power amplifier stages.

The main advantage of solid-state circuitry in high power AM broadcast transmitters is operating efficiency. It is still yet to be determined if the reliability of solid-state designs is distinctly advantageous over power grid vacuum tube circuitry, but the efficiency advantages are clear and proven. The efficiency of a single power grid vacuum tube pulse width modulator, including filament, control grid, and screen grid power losses is approximately 90% maximum for a 50 kW transmitter. Pulse-step/width modulator efficiency for a solid-state modulator at the same power level is approximately 95% or higher. When solid-state circuitry is also employed in the RF amplifier circuitry, the efficiency advantages are even greater. The typical maximum anode efficiency of a power grid vacuum tube Class-C RF power amplifier is 84% neglecting output network circuit losses. This efficiency can be improved to approximately 90% by anode voltage wave-shaping techniques such as 3rd and 5th harmonic traps in the output anode network. Therefore, the all tube pulse-width modulated transmitter overall modulator/RF power amplifier (RFPA) anode efficiency is 90%*90%, or approximately 81%. Transmitter designs employing solid-state circuitry in the RF power amplifiers typically use Class-D RF power amplifier circuitry which achieve approximately 95% RF collector efficiency at medium wave frequencies. Therefore, an all solid-state pulse-width modulated transmitter can have overall collector efficiency of 95%*95% = approximately 90%, 9% greater than the all tube transmitter. With energy costs constantly rising, the efficiency of broadcast transmitters will continue to be very important.

Thomcast Pulse-Step Modulator

Thomcast of France and Switzerland was one of the first companies to market a workable very high power solid-state modulator for use with their family of LF, MF, and SW transmitter products, from 100 to 1,200 kW of carrier power. The system employs several low voltage power supplies in series which are switched into operation in steps to provide the required anode modulation voltage, hence the term *step modulation*. The concept was first employed in a 500 kW shortwave transmitter design and later scaled down at LF, and MF, as well as other shortwave products of lower power levels to 100 kW, minimum. Typically the number of steps employed is approximately 28. Each step is pulse-width modulated to achieve the necessary total linearity required for broadcast transmitter modulation linearity characteristics. Each step represents approximately 1,000 V in modulator output, yielding a total modulator output capability of approximately 28,000 V at the 100% positive modulation crest.

The transmitters were designed to comply with the ITU international shortwave broadcast standard of 4.5 kHz audio bandwidth which is also compatible with the requirements for audio bandwidth of ITU region 2 and 3 medium and long wave transmissions.

To achieve high overall transmitter efficiency, Thomcast engineers concentrated not only on the solid-state modulator circuitry, but also on techniques to maximize the anode efficiency of the power grid vacuum tube final RF power amplifier and associated circuitry. Working closely with the power grid vacuum tube division of Asea Brown Boveri (ABB), they designed a tube with optimum circuit efficiency characteristics which operates with minimum grid drive power requirements, an anode conduction angle of approximately 110°, rather than the more conventional 120° standard for Class-C operation, yielding a total anode efficiency of approximately 87%, excluding circuit losses. Medium wave transmitter designs by Thomcast using pulse-step modulation employ similar technologies with the addition of anode wave-shaping circuitry, resulting in anode efficiencies exceeding approximately 92%. Overall, shortwave transmitter efficiency has been measured at greater than 76% on some international shortwave bands. Overall efficiency of Thomcast medium wave transmitters has been measured at greater than 77%.

Continental Electronics Corporation (CEC) Pulse-Step Modulator

Continental Electronics Corporation (CEC) introduced their own version of a high power pulse-step modulator system in 1990. The system is similar to the ABB Thomcast system in basic operating principle. The technique used to generate each step and the subsequent pulse-width modulation of each step is different than the system employed by Thomcast, but the net result is similar.

The CEC system for the basic step generation uses a modified flash A/D converter with an additional digital ring modulator system that electrically rotates the active steps equally among all available modules. This rotation provides equal power distribution among all system modules. Each step is pulse-width modulated by a linear mixing of a triangular dither signal to each A/D flash comparator.

The CEC system employs typically approximately 48 discrete steps of approximately 750 V each, and has been built for transmitter carrier power levels of 100 kW, 250 kW, 500 kW and higher, both medium wave and shortwave. The CEC design employs high speed IGBT insulated-gate-bipolar-transistors (IGBT) producing typical modulator efficiencies of 95% or higher, overall transmitter efficiencies of greater than 72%, and modulation bandwidth capabilities exceeding 10 kHz.

Nautel Corporation—AMPFET Transmitters

The Nautel Corporation of Nova Scotia, Canada was one of the first to introduce all solid-state transmitter designs on the commercial broadcast transmitter market. Initially available for only 1, 5, and 10 kW carrier powers, they have since added capability to 50 kW and to 100 kW by combining two 50 kW transmitters.

The Nautel design employs independent low power modules, each with Class-D RF power amplifiers with pulse-width modulation of the Class-D amplifier collector voltage. Each module contains its own RF power amplifier and pulse-width modulator and approximately 24 modules are then combined in an RF combining network to achieve the desired carrier power level. The collector efficiency of each Class-D RF Power amplifier (RFPA) is approximately 95% and the efficiency of the combining networks is greater than 99%, yielding a total RF system efficiency of approximately 94%. The efficiency of the pulse-width modulators is approximately 95% resulting in an overall collector efficiency of better than 89%. Nautel engineers then carefully selected and designed the remaining transmitter components to require no more than approximately 10% of nominal carrier power for all auxiliary transmitter functions, such as control circuitry, exciter and driver circuitry, cooling equipment, etc, thus achieving an overall transmitter efficiency of greater than 80%. The basic circuit concept of the Nautel design is shown in Figure 4.1-10.

Harris Broadcast Division, Polyphase Pulse-Width Modulation

In the late 1960s, Harris Corporation developed their version of pulse-width modulation—Pulse Duration Modulation (PDM). In the original configuration, audio information to be transmitted is sampled at a nominal 75 kHz rate. This creates a 75 kHz pulse train with its pulse width rate-of-change being equal to the audio frequency. Studies have determined that the time duration of the sample pulse is not critical. The sample rate is based on an upper frequency of 15 kHz sampled at five times that frequency to ensure broadcast quality bandwidth and distortion. In practice, most transmitters manufactured for use in North America employ a switching frequency of approximately 70 kHz, to simplify compliance with the FCC requirement for suppression of spurious emissions beyond 75 kHz from the carrier frequency.

The technique of polyphase pulse-width modulation (PPDM) was a further development to overcome the switching transition losses of the higher sampling frequency, the effective sample rate being four times normal, or approximately 280 kHz.

When extracting the long and short term dc average information from an encoded single phase PDM signal, a sharp filter roll off characteristic is required. This sharp attenuation characteristic generates overshoot due to the inherent group delay characteristic of the sharp roll off filter. The effective 280 kHz sampling rate of the Harris four phase sampling system, however, allows some relaxation in filter cutoff requirements that significantly reduce overshoots.

Harris Broadcast Division, DX Series Pulse-Code Modulation

One of the most exciting and truly innovative improvements to the AM broadcast transmitter art is the Harris Broadcast Division DX series of medium wave AM broadcast transmitters employing direct digital techniques to generate the amplitude modulated wave. The technique has no modulator as such and the complete AM wave, carrier and modulation sidebands, are produced by direct digital pulse-code control of several small power solid-state RF amplifier modules.

The DX system is employed in transmitters ranging in power from 5 to 50 kW for the domestic U.S. market, and has been sold at 1,000 kW power levels for the ITU Region 2 & 3 markets for which carrier powers up to 1 and 2 MW are not uncommon.

The system has many inherent advantages over conventional AM modulation systems.

- It is solid-state and employs no separate modulation of the final RF amplifiers (the amplitude modulated wave is directly achieved by digitally selecting series connected small RF output modules)
- Having no audio modulator, it is capable of high quality modulated wave characteristics such as lower modulation noise, distortion, and transient modulation characteristics
- Being basically a digital system, it is capable of directly receiving digital audio input from any variety of digital audio sources, such as CD's, RDAT, digital satellite program feeds, or future ISDN program distribution, without D/A conversion.

The Harris DX system is shown in Figure 4.1-11. The incoming analog signal is converted to digital code by conventional sample-and-hold and pulse-code modulation (PCM) techniques and circuitry. The top 6 (most

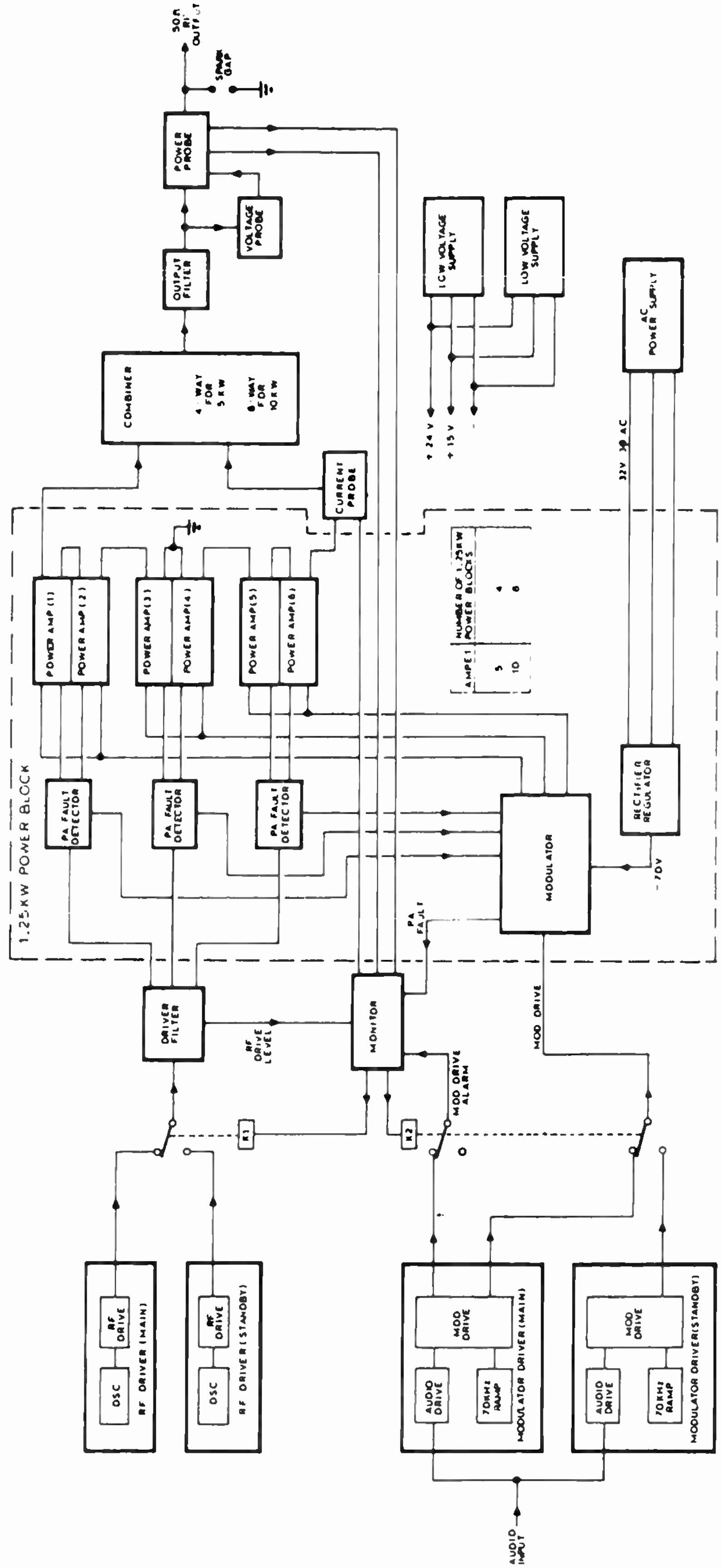


Figure 4.1-10. Block diagram, Nautel Model AMPFET 5 and AMPFET 10, 5 & 10 kW AM broadcast transmitters.

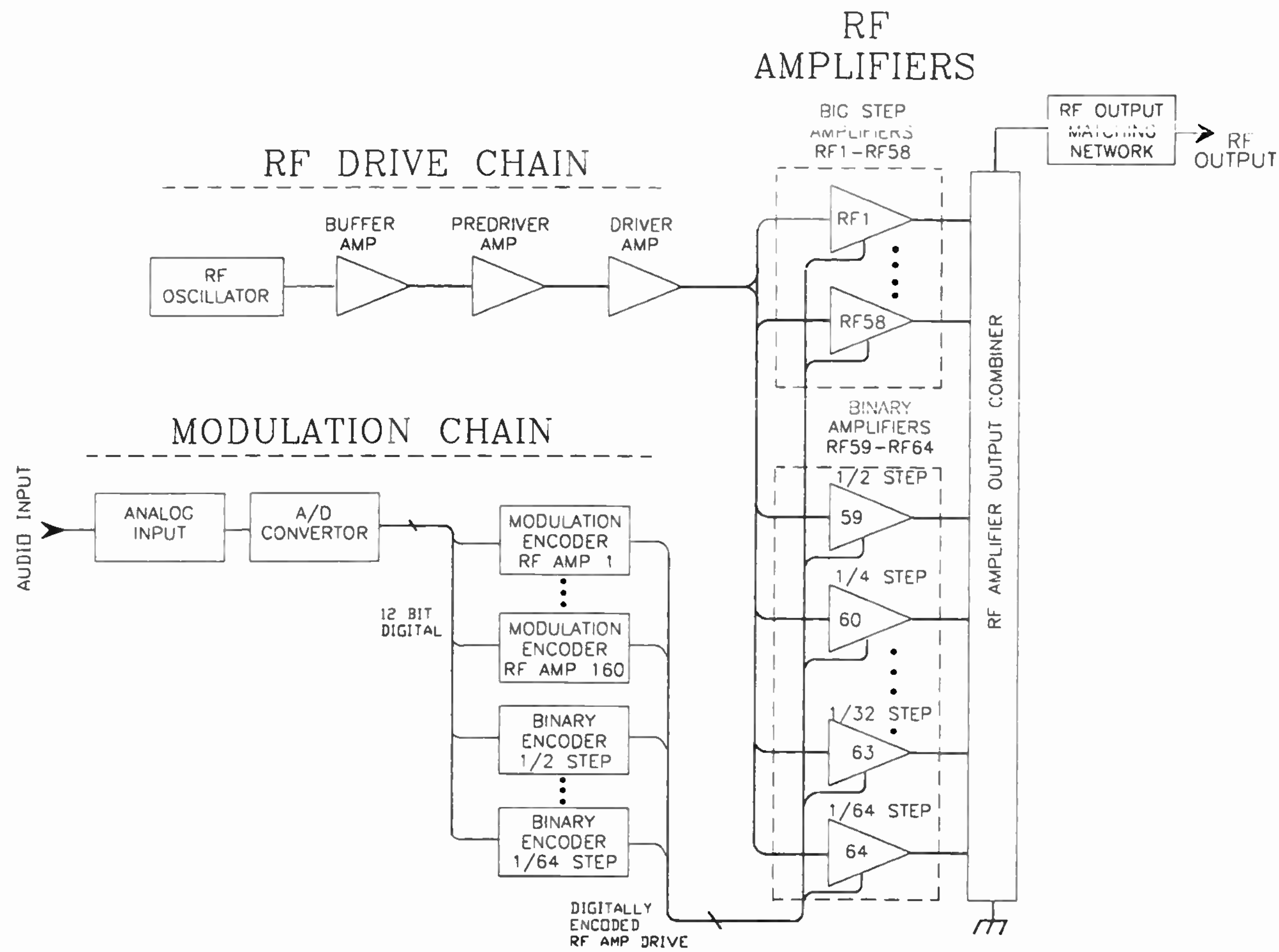


Figure 4.1-11. Harris DX system.

significant) bits of a twelve bit A/D PCM system are converted into 64 equal steps. The top 22 steps are not used leaving 42 equal steps which are used to control 42 equal output series RF amplifier modules. This is equivalent to keeping the lower 5.4 of the 6 top bits of the PCM digital code ($2^{5.4} \approx 42$). The bottom 6 (least significant) bits are then used to control 6 half-step RF amplifiers (amplifiers with outputs of 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64th of the major 42 equal step amplifiers). In effect, this creates a digital reconstruction of the analog input waveform of 2709 discrete steps ($(42 + 1)*63 = 2709$). This is roughly equivalent to a PCM system with a total resolution of 11.4 bits ($2^{11.4} \approx 2702$). The theoretical capabilities of such a PCM system are well known yielding a theoretical harmonic distortion value of approximately 0.03% and audio noise of less than -70 dB referred to 100% modulation.[13] Typical measured values of THD are approximately 10 to 30 times this theoretical figure due to the practical difficulty in achieving balanced output contribution from all of the series output modules. However, even these levels of harmonic distortion (less than 1% maximum) are well below accepted norms for modern high power AM broadcast transmitters. Typical values of measured audio noise are typically very close to the theoretical maximum of -70 dB referred to 100% tone modulation at full carrier power.

Each high power amplifier module is a solid-state amplifier operating in the Class "D" mode having approximately 95% power efficiency, including output circuit losses. Since there is no additional modulator losses; (the modulation process takes place in the RF circuitry itself) the auxiliary circuits of cooling, control, excitation and drive, etc, represent the only other input power requirements. The total auxiliary input power of a 50 kW transmitter is approximately 5 kW. This means that the total input power of a 50 kW transmitter at carrier only power (no modulation) is approximately: $50/0.95 + 5 = 57.63$ kW, yielding an overall efficiency of approximately $(50/57.63)*100\% = 86.76\%$. Harris Broadcast Division claims 87% or greater overall efficiency for all their DX series transmitters greater than 5 kW carrier power, truly a remarkable achievement.

TRANSMITTER CIRCUITRY

Detailed transmitter circuit design is as varied as the individual designers. There is, however, some basic circuitry which is common to all transmitter types and models that will be briefly discussed in this section.

Carrier Frequency Generator/Exciter

The quartz crystal oscillator has been the stable frequency source for all transmitters manufactured since about 1930. The quartz crystals used in older model transmitters were large cuts of natural quartz, vacuum sealed in glass envelopes, similar to small power vacuum tube envelopes, and occasionally mounted in temperature controlled ovens to obtain the required FCC carrier frequency stability. It is more common in modern designs for the quartz crystals to be enclosed in small hermetic sealed metal cans made popular and proven reliable by the military and commercial communications equipment industry, and without special temperature controlled circuitry. Modern quartz crystal manufacturing technology and the use of proven solid-state crystal oscillator designs allows these types of small metal sealed units to adequately maintain the FCC's current requirement of ±20 Hz carrier frequency tolerance. Stability of the quartz oscillator circuits is normally adequate over the full lifetime of the equipment. Should frequency adjustment ever be required to bring a unit back inside the FCC limits, mechanically stable adjustment components, usually a glass or ceramic piston type of capacitor, are provided for use by a qualified station engineer using calibrated frequency measuring equipment. (See FCC Rules and Regulations Sections 73.1540 and 73.1545). Exciters for proposed AM stereos provide the carrier frequency excitation for the transmitter as well as the stereo generating circuitry. Because of manufacturing advantages, some manufacturers of AM stereo exciters generate the desired carrier signal with frequency synthesizer techniques. This method of generating the basic carrier frequency is generally equal to or better than the discrete quartz oscillator method with regard to frequency tolerance but can produce higher phase modulation noise if improperly designed or adjusted.

RF Power Amplifier

As discussed in earlier sections, all modern AM broadcast transmitters using power grid vacuum tube final RF amplifiers employ Class-C amplification. Some employ designs using third and fifth harmonic trap circuitry yielding quasi-class "D" operation for improved efficiency. All solid state transmitters use some form of Class "D" RF final power amplification.

RF Output Networks

The purpose of the RF output network is to match the impedance of the load (the common point impedance of one or more antenna matching and combining networks) to the impedance required by the final RF power amplifier tube or transistor to produce the desired carrier and sideband power. The output network circuit is also designed to provide the attenuation characteristics necessary to meet the FCC's requirements for spurious and harmonic output. There are many techniques to accomplish these basic tasks. One simple and effective method is shown in Figure 4.1-12. The terminating load impedance for the network shown in Figure 4.1-12 is defined by the Smith chart representation of Figure 4.1-13, which is a typical common point impedance characteristic of a multitower AM broadcast directional antenna array. Figure 4.1-14 shows the impedance versus frequency characteristics at the input of the network; at the anode of the final RF amplifier tube. The shape of the impedance curve in Figure 4.1-14 differs significantly from the shape of the terminating impedance curve shown in Figure 4.1-13 due to the narrowing of bandwidth caused by the output matching network. The VSWR (normalized to the resistive carrier impedance) at ±10 kHz is about 1.2 in both Figures 4.1-13 and 4.1-14 but at ±50 kHz the VSWR at the input to the network has increased from about 2.4 to approximately 3.0. The shape of the impedance versus frequency curve at the anode of the RF output amplifier tube (or transistor collector for the case of solid-state output devices) has an important role to play in the high audio frequency performance characteristics of the transmitter, such as frequency response and harmonic distortion.[14] The shape of the impedance versus frequency curve should be symmetrical about the resistive axis of the Smith Chart and should yield lowest practical VSWR values at the highest expected fundamental Sideband frequencies. Figure 4.1-15 shows the result of an attempt to lessen the magnitude of the mismatch at the input to the matching network by the addition of the dotted components shown in Figure 4.1-12. The result of adding the dotted components in Figure 4.1-12 is a significant reduction in tube anode load mismatch at sideband frequencies of ±10 kHz (from about 1.2 VSWR in Figure 4.1-14 to approximately 1.1 VSWR in Figure 4.1-15). The reduction of mismatch at ±10 kHz, however, comes at the expense of a worsening of mismatch at sideband frequencies greater than ±30 kHz. The worsening mismatch at higher sideband frequencies is normally not a problem and often tends to improve transmitter harmonic distortion performance, acting as an output bandpass filter to higher order harmonics. There is theoretically no limit to the amount of sideband mismatch correction that can be achieved with more complex network circuitry. The most important criteria is symmetry about the Smith Chart resistive axis. Station engineers or their engineering consultants should work with the transmitter manufacturers and antenna system designers to, first determine, then achieve the desired impedance versus frequency curve at the output RF amplifier tube anode connection. The example given for load sideband mismatch correction is for illustrative purposes only. It is quite common for the complex impedance at the sideband frequencies to present a more severe mismatch at the antenna common point

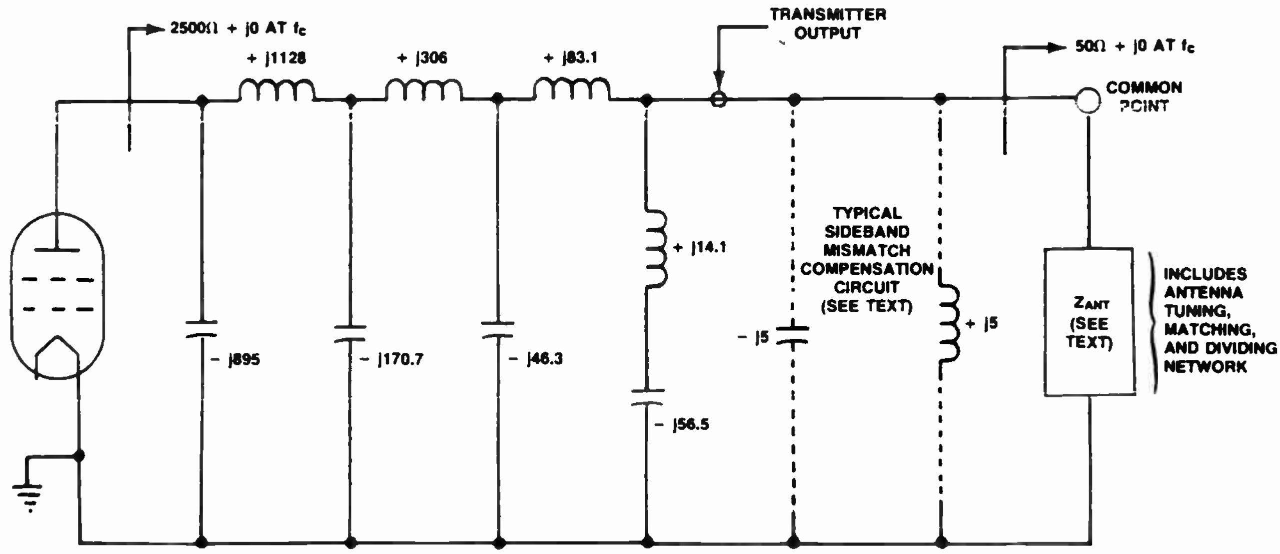


Figure 4.1-12. Typical output matching network yielding 80 dB or better harmonic attenuation and a compensated symmetrical sideband load at the anode(s) of the RF output amplifier.

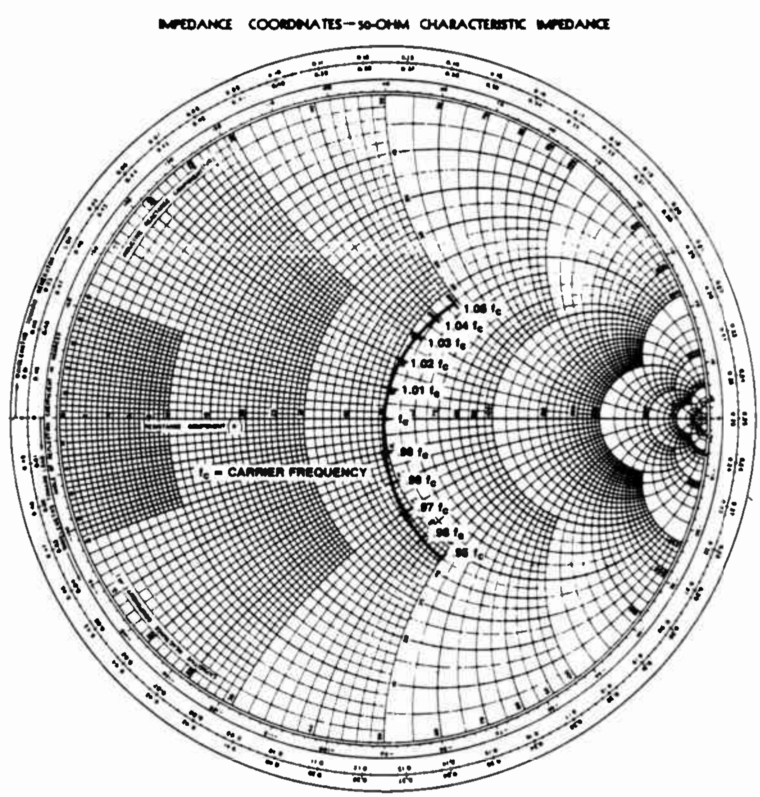


Figure 4.1-13. Typical impedance versus frequency characteristic of antenna common point.

IMPEDANCE OR ADMITTANCE COORDINATES

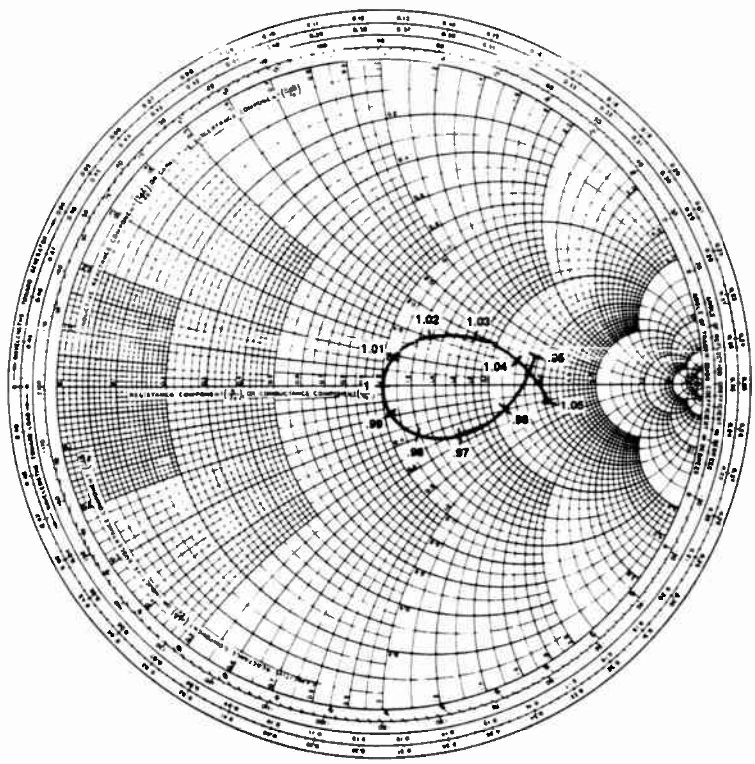


Figure 4.1-14. Impedance versus frequency characteristic at anode of final RF power amplifier.

IMPEDANCE OR ADMITTANCE COORDINATES

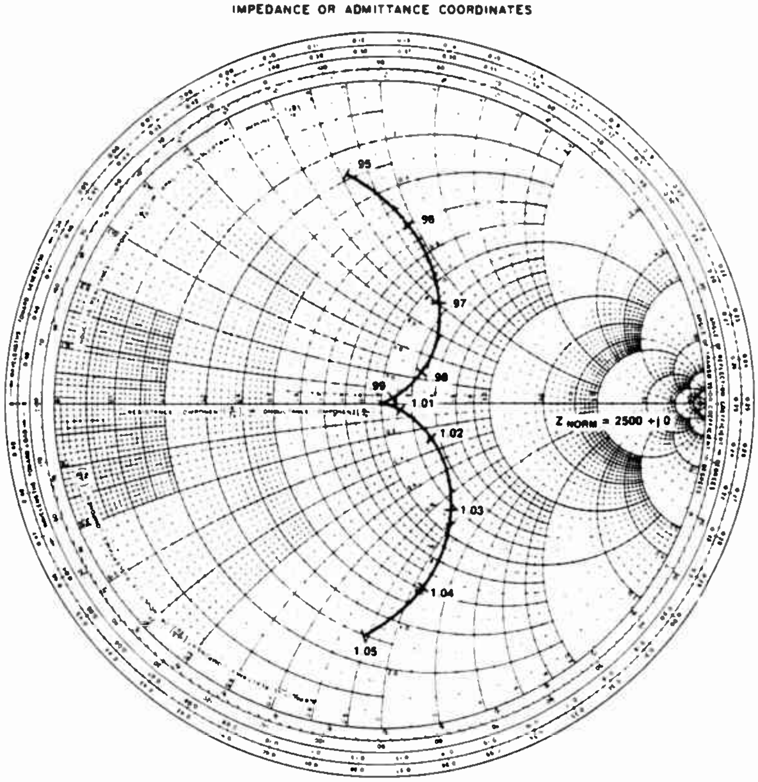


Figure 4.1-15. Impedance versus frequency characteristic at output anode with sideband mismatch corrective network added.

than that shown in Figure 4.1-13. More severe mismatch and impedance dissymmetry requires more complex sideband mismatch corrective networking.

Transmitter Control and Monitoring

It is common to find the task of transmitter control performed with discrete digital IC logic circuitry in modern designs that used to be accomplished with simple relay control logic. Some manufacturers are incorporating microprocessor technology in their latest equipment designs to replace discrete digital logic circuitry. The aim of manufacturers is normally to provide the operating engineers and technicians with the most basic, reliable and easy to maintain and troubleshoot transmitter possible. Experience has shown that well designed relay control logic, discrete digital IC logic and microprocessor based logic are all about equal in terms of reliability and ability to perform the required basic transmitter control functions. Future microprocessor based transmitter control logic offers the promise to provide self and remote assisted diagnostics of transmitter problems and remote interrogation of transmitter operating parameters. Microprocessor based logic offers the added promise of altering the basic characteristics of a transmitter control system through software control, allowing the basic transmitter design to be more easily customized to individual users' operating requirements. Because of the high voltage and high current faults that can exist in any high power transmitter component, extra care must be taken by designers and manufacturers of high power AM broadcast transmitters to prevent the potential destructive energy in these faults from affecting the performance and operation of the relatively delicate solid-state control logic circuitry. High speed vacuum contactors and solid-state regulator/controllers are being used in many modern designs to control the high voltages and currents encountered in all levels of high power AM transmitters, which previously had been controlled by slower though equally reliable air-magnetic contactors and relays.

Remote control systems are commercially available that will allow remote control of almost any transmitter, old or new. These remote control systems, like the basic transmitter control system, use relay, discrete digital IC, microprocessor based logic circuitry, or combinations of these, depending on the manufacturer of the remote control equipment and the complexity of the remote control functions desired. Many transmitter equipment manufacturers provide limited built-in remote control functions and circuitry.

The FCC requires manufacturers of AM broadcast transmitters to provide certain basic monitoring functions for proof of certain performance parameters. These requirements are contained in the FCC Rules and Regulations Section 73.1215.

High and Low Voltage Direct Current Power Supplies

Direct current power supplies in AM broadcast transmitters must be designed to provide minimum acceptable performance in two basic areas: power supply ripple, which affects transmitter hum and noise output, and dynamic regulation, which affects low frequency modulation transient response. It is typical for transmitters of 5 kW carrier power and lower to operate from single phase ac power sources, usually 240 V in North America. Transmitters with carrier power ratings of 10 kW operate with single-phase or three-phase power sources, depending upon the manufacturer of the transmitter. Transmitters with carrier power ratings of greater than 10 kW operate only from three phase power sources, usually 480 V for power levels up to 100 kW and 4160 or 11,000 V in North America for higher carrier power levels. Three-phase power has the advantage of being easier to filter and usually provides better dynamic regulation of critical modulator voltages than single phase supplies. Single-phase power is more readily available which is the only reason it is used at the lower transmitter power levels, because initial installation cost would be disproportionately increased were three-phase power required. Single-phase rectifier power supply systems generally require L/C filtering to provide the necessary low ripple output for low transmitter hum and noise specifications. L/C filtering also creates, however, power supply resonances in the audible to sub-audible range of modulating frequencies and therefore is a source of poor dynamic power supply regulation when the modulator circuitry of the transmitter is excited by vowel sounds or musical percussion sounds. It has been common since about 1970 for higher power transmitters to use special high voltage supply transformers to generate a six-phase ac supply from the basic three-phase power source. The six-phase supply, when full wave rectified, yields a 12 pulse rectified dc waveform that has both lower ripple content and a higher ripple frequency than conventional three-phase full wave rectification. As a result, the output of the rectifier can be sufficiently filtered with no additional filter inductors, thus improving low audio modulating frequency dynamic power supply regulation and hence, low frequency transient distortion.

TRANSMITTER PERFORMANCE MEASUREMENTS

Certain basic AM transmitter performance parameters should be measured on a periodic basis in order to assure that certain minimum broadcast quality standards are provided to the listening public. Excluding the performance standards for AM stereo, these are listed below. Some are specifically required by the FCC, while others are a matter of good engineering practice.

a) Spurious emissions (Section 73.44). This measurement must be made annually, with no more than 14 months between readings (Section 73.1590(a) (6)).
b) Operating power (Part 73.51)
c) Carrier output power delivered to the antenna system (Part 73.54)
d) Modulation capability

e) Total audio frequency distortion
f) System frequency response
g) Carrier-amplitude regulation (Carrier Shift)
h) Hum and noise output level
i) Carrier frequency tolerance [Part 73.1545(a)]

All audio measurements should be made from a demodulated voltage or current sample at the antenna system common point.

Regarding item c, the measurement of transmitter power output by the direct method described in Parts 73.51 and 73.1215 is subject to more than 13% error if allowed FCC measurement inaccuracies are taken to the limit. For example, assume that a common point impedance of 50 Ω resistive can be measured to within 2% accuracy with a radio frequency impedance bridge, a realistic tolerance. Further assume a direct reading RF ammeter having a full scale reading of 100 A and an FCC allowed tolerance of ±2% of full scale is used for this measurement and indicates 33.33 A of common point current (just meeting the minimum FCC accuracy and indication requirements in Part 73.1215(b)(2) and (3)). Under these conditions, the actual power delivered to the antenna is between the limits of (35.33 × 35.33 × 51) 63.65 kW, as a maximum, and (31.33 × 31.33 × 49) 48.10 kW, as a minimum; yielding a total measurement error of approximately +14/−13%. Using an RF ammeter with a 50 A full scale reading, the same FCC allowed inaccuracies, and the same meter indications as above, would result in power output measurement errors of approximately ±8%.

FACTORY TESTS

When looking for an AM broadcast transmitter to buy for replacement or new equipment, it is advisable to ask manufacturers for specific and detailed test and performance data at the start of the buy investigation. Before a final decision is made, specific and detailed tests should be made at the manufacturer's factory under strict control of an experienced engineer or engineering consultant. Most manufacturers of AM broadcast equipment welcome this kind of intelligent approach to the decision making process. Following are some hints on what kind of tests and a discussion on the details of each test.

Audio Frequency Response

In the early days of radio broadcasting, the term *broadcast quality audio* meant a standard to which all other system equipment, by comparison, was inferior. Today the audio quality of consumer high fidelity and stereo equipment surpasses, with one exception, the signal quality which can be broadcast by any AM transmitter manufactured today or likely to be made in the future. The one exception is audio frequency response. Practically all mass production AM receivers made in any country of the world have IF and audio amplifier bandpass characteristics that limit receiver −3 dB high end audio frequency response to between approximately 2,500 Hz and 5,000 Hz, with 2,500 Hz the more common of the two figures. (This poor high frequency response of the receiver is not a limitation of technical capability, rather a conscious decision on the part of receiver manufacturers to limit high end response in order to minimize perceived nighttime skywave co-channel and adjacent channel interference on the medium-frequency AM band.) Typical low end −3 dB frequency response of consumer AM radios is between 100 Hz and 300 Hz, with approximately 200 Hz a common value. AM transmitter frequency response characteristics, referred to a reference frequency of either 400 Hz or 1,000 Hz, typically are well within ±1 dB from 50 to 10,000 Hz at any depth of modulation, exclusive of antenna system characteristics which can and often do have a deleterious effect on many audio performance system characteristics.

Audio Harmonic and Intermodulation Distortion

Current production AM broadcast transmitters typically produce less than 2% (THD) up to 90% modulation at any frequency of modulation between 50 and 10,000 Hz for monophonic transmission. Most of the modern all solid-state digital systems produce typical audio distortion of less than 1% THD at most conditions of modulation.

Intermodulation distortion (IMD) has been known for years to be a more disturbing kind of distortion than harmonic distortion, though both are important. The ITU-R method of IMD measurement is the preferred method for radio broadcast transmitters. With this method, two equal audio tones separated by approximately 170 Hz are input to the transmitter and the peak modulation level adjusted to between 85 and 95% modulation. The level of odd and even order products are measured using an audio wave or spectrum analyzer connected to the test output terminals of a high quality modulation monitor. Two IMD measurements should be taken, one with the two tones near mid audio band (for example: 400 and 570 Hz) and one with the two tones near the upper audio end (7,000 and 7,170 Hz). High quality broadcast transmitters should produce odd order IM distortion products more than 30 dB below the level of either of the two modulating tones. The RSS value of all IM products, relative to the level of either modulating tone, should also be less than 5% at 90% peak modulation levels.

Residual AM Hum and Noise

The FCC requirement for AM noise of about 60 dB below the 400 Hz/100% modulation level can be achieved by most current production AM broadcast transmitters. The bandwidth of noise measuring equipment should be 20 kHz minimum. Typical modulation monitor demodulated audio bandwidth is approximately 25 kHz to accommodate both of these FCC bandwidth requirements.

Residual PM Hum and Noise

Residual phase modulation (PM) noise is normally not a problem for modern AM broadcast transmitter designs for monophonic broadcasting. The quartz crystal oscillator circuits and even moderately careful RF component mechanical designs produce quite acceptable phase noise characteristics. However, for AM stereo applications, where a frequency synthesizer may be used for the RF signal generation, is it wise to perform a test to determine the purity of the synthesizer circuitry. Very excessive residual PM noise can convert to AM noise over certain nighttime propagation paths and appear to distant listeners as objectionable AM hum and noise. An acceptable value of PM noise is −25 dB rms relative to one radian peak, measured in a 15 kHz bandwidth, for monophonic medium wave band AM transmission. Transmitters with quartz crystal oscillator exciters typically exceed this recommendation by 25 dB or more. Transmitters used for international shortwave broadcasting (4 to 26 MHz) require PM noise levels of approximately −45 dB relative to a one radian peak because of more severe sky-wave PM to AM conversion at higher frequencies.

Incidental Phase Modulation (IPM)

Like residual PM, IPM is more important in stations using or planning to use AM stereo transmission. IPM is defined as the peak phase deviation of the carrier frequency (in radians) resulting from the process of AM. IPM values of several radians were common in the very early days of broadcasting. Typical values of IPM for modern transmitters that have not been specifically designed or adjusted to minimize IPM range from about 0.1 to 0.5 radians peak (approx. 6 to 30°). A maximum acceptable value of IPM required for AM stereo operation is generally considered to be approximately 0.05 peak radians with a desired value of better than 0.02 peak radians. State-of-the-art modulation meters such as the Hewlett Packard model HP-8901B and other similar instruments are preferred for accurate PM and IPM measurements.

Carrier Amplitude Regulation (Carrier Shift)

The amount of carrier level shift in a given transmitter has more importance to overall transmitter performance than many broadcast engineers realize. Large values of negative carrier shift can have as much effect on effective transmitted sideband power as poor transient overshoot distortion. The term *carrier shift* may be somewhat confusing, especially to newcomers in radio broadcasting who often equate the terminology with frequency shift instead of level or amplitude shift, hence, the new terminology of *carrier amplitude regulation*. The ITU-R refers to the same characteristic as *carrier level shift*, which appeals to many engineers because of its closer adherence to the original terminology, but without the ambiguities. Carrier level shift is the effective shift in apparent carrier level due to the AM process. Carrier level shift can be caused by either poor power supply regulation, or modulation even order harmonic distortion, which generates a dc offset component in the modulated RF envelope, or both. Carrier level shift can be either positive or negative, though is usually negative because power supply regulation is most often the major source of carrier level shift and power supply regulation is most generally negative in sign (lower voltage output at higher current loads). Poor power supply regulation is not always caused by the transmitter power circuitry, it can also be, and often is, caused by poor supply line voltage regulation or, more generally a combination of the two.

Another common misconception regarding carrier level shift is that the defined level shift is a direct shift in carrier power. Actually, carrier level shift is defined as the shift in effective carrier *voltage or current* due to the process of modulation. This means that a carrier level shift of minus 5% is equivalent to a carrier power shift of approximately minus 10%.

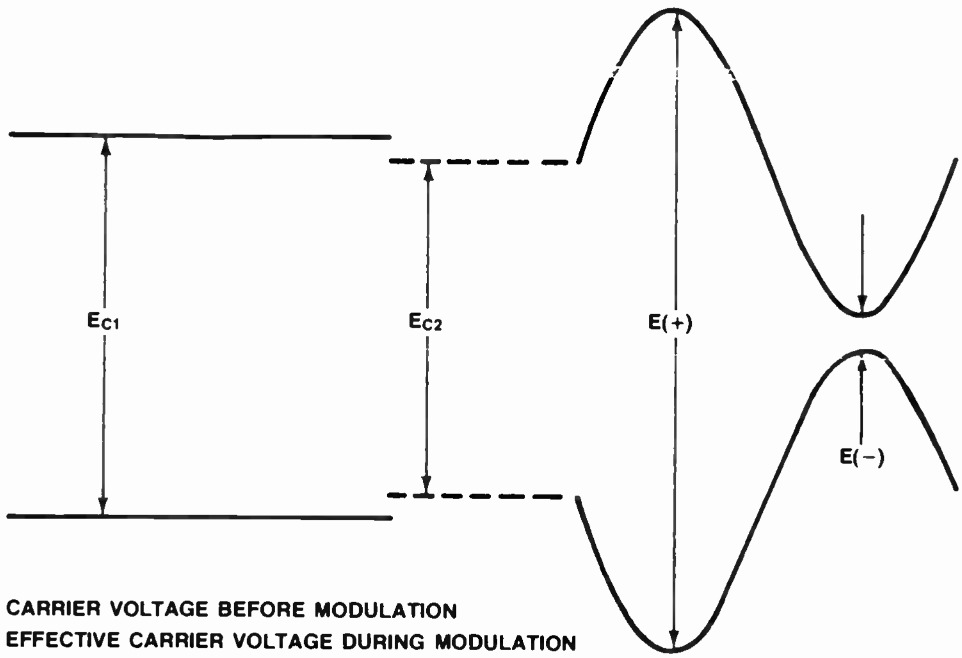
$$\begin{aligned} P_{carrier\,(mod)} &= P_{carrier} \times (1 - 0.05)^2 \\ &= 0.9025 \times P_{carrier} \end{aligned}$$

Transmitters having no carrier level shift produce an average output power of 1.5 times the carrier power level with 100% sinusoidal tone modulation. Transmitters that exhibit −5% carrier level shift produce an average output power only 1.35 times the carrier level at the same conditions of symmetrical sinusoidal tone modulation. A broadcast station engineer or engineering consultant needs to have complete understanding of these and other equally important transmitter characteristics in order to be able to make an intelligent buying decision based upon measurable and proven technical merit. Figure 4.1-16 gives a graphical representation of carrier level shift.

Audio Phase Linearity

Proper attention is given to phase linearity by most station engineers and engineering consultants. Station managers and engineers are usually properly concerned about the sound of their stations; and loudness, or perceived loudness, is a common criteria of quality in many stations with diverse programming formats. However, it is not uncommon to learn that many station engineers and program directors spend more time researching the various objective and subjective merits of program processing equipment than the one characteristic in their potentially more expensive transmitter purchase that could partially neutralize the potential advantage from a new or different program processor—that one characteristic is audio phase linearity.

Audio phase non-linearity and its major detrimental result, transient overshoot, have been discussed in previous sections of this chapter. Before the popularity of modern broadcast audio processing and some modern broadcast programming formats, audio phase linearity of a transmitter or any component in the program audio chain, was not as much of a concern as it is today. (Modern programming philosophy, and the resulting audio processors have made transmitter audio phase linearity characteristics a more important performance criteria.) In the late 1960s, at the long and persistent insistence of the AM broadcast industry, the FCC reluctantly authorized 125% positive program modula-



E_{C1} = CARRIER VOLTAGE BEFORE MODULATION
E_{C2} = EFFECTIVE CARRIER VOLTAGE DURING MODULATION

$$\text{CARRIER LEVEL SHIFT (CS)} = \frac{\frac{E(+) + E(-)}{2}}{E_{C1}}$$

$$\text{\% MODULATION} = \frac{E(+) - E(-)}{E(+) + E(-)} \times 100$$

Figure 4.1-16. Graphical representation of carrier level shift and formulas for calculation of carrier level shift and % modulation.

tion, allowing AM transmissions to accommodate certain naturally occurring asymmetry in voice and music thus achieving a gain of 2 dB of real loudness or actual program sideband power (20 log(1.25) = 2 dB). As stated earlier, some AM broadcast transmitters in present use exhibit as much as 12% overshoot of a squarewave input, due to phase non-linearity, which has the effect of taking away one of those 2 dB the industry tried so hard to get (20 log(1/1.12) = −1 dB).

A simple way to determine the effects of phase non-linearity, is to require the prospective transmitter manufacturer to demonstrate rectangular wave modulation characteristics of the transmitter under investigation. When such a test is performed, the overshoot produced by the transmitter will be directly visible and measurable on the oscillographic display of the transmitter output envelope. The station engineer or his engineering consultant should also investigate the effects that antenna system phase non-linearity may have on the total system transmission transient characteristics.

Occupied Bandwidth

Occupied bandwidth of an AM broadcast transmitter can best be measured with the use of a band limited colored gaussian noise source, similar to pink noise used in certain acoustical tests, to provide a continuous wideband modulating signal to the transmitter. Pink noise has an equal energy bandwidth ratio (equal energy per octave, per third octave, per tenth octave, etc.). White noise has equal energy per bandwidth (equal energy per Hz). Both noise signals can have a Gaussian or pseudo-Gaussian probability density function, a probability density function that closely resembles that of typical voice and music. Figure 4.1-17 shows a block diagram of the test procedure used to measure occupied bandwidth. (See CCIR recommendation 559-1 (Vol X Part 1 Geneva, 1982). The measurement of occupied bandwidth is a dynamic measurement which effectively summarizes two important transmitter static parameters: audio non-linear distortions, which are the source of IMD and THD, and IPM.

Harmonic and Spurious Output

Harmonic and spurious output of an AM broadcast transmitter or transmission system can only be effectively measured two ways. One way is to use a sample of the transmitter RF output signal when operating into a dummy load with a known or accurately measurable impedance characteristic out to approximately the tenth RF harmonic. Then, with a calibrated measuring system, measurements and calculations are made to

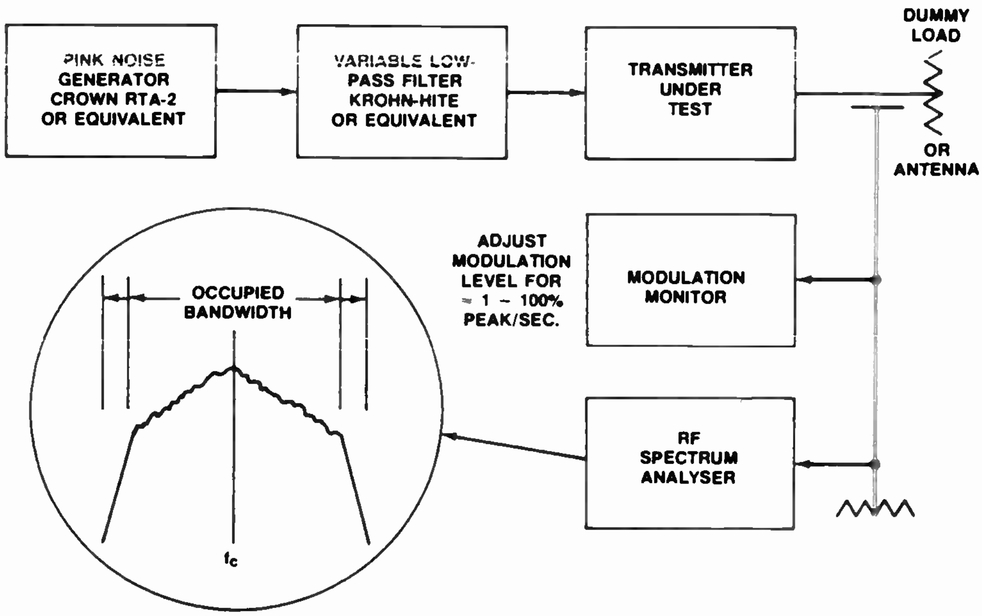


Figure 4.1-17. Block diagram test procedure for determining transmitter occupied bandwidth.

compute the output power of each harmonic according to the formula:

$$P_n = V_n^2 \times R_{pn} \qquad [8]$$

Where:

P_n = the power at the Nth harmonic
V_n = the corrected measured voltage of the Nth harmonic at the calibrated impedance point
R_{pn} = The parallel resistive component of the load impedance at the Nth harmonic

A second way to effectively determine the harmonic and spurious output of an AM broadcast transmitter is to measure the actual power radiated from the antenna system at each harmonic frequency or suspected spurious output frequency by standard field intensity measurement techniques. The field intensity method is the most meaningful of the two techniques because it allows the measurement to be made under actual conditions of operation with all systems interconnected. It is normally the new transmitter manufacturer and users joint responsibility to correct any actual interference problems to other broadcast or non-broadcast communication services even when the particular interfering signal meets the standard FCC requirements (see FCC Rules Section 73.44.(c)).

Carrier Output Power

The most accurate method of measuring the RF output power of a transmitter is by the calorimetric method, a method which uses the very accurately known and measurable physical and thermal characteristics of water or other similarly well defined liquids. This measurement is usually only done in a transmitter manufacturer's factory because the capital investment required to purchase and maintain calibration of this kind of infrequently used equipment is usually not justified for AM broadcasting operations.

Water is known to have a thermal capacity very close to 4.186 Joules per degree C per gram weight at a mean temperature of 60° centigrade. A Joule is equal to one watt-second. Therefore, the capacity of water to absorb power is 69.8 watts per degree C per liter of water flow per minute, or

$$\text{Power (kW)} = \text{Flow (lpm)}*\text{DT°C}*\ 0.0698$$

for water flow measured in liters per minute (lpm)

or,

$$\text{Power (kW)} = \text{Flow (gpm)} \times \text{DT°C} \times 0.2641$$

for water flow measured in U.S. gallons per minute (gpm)

The flow of water can be measured with an accuracy of approximately ±1%, by even the most common methods. Differential temperature measurement accuracy of approximately ±0.1°C is commonly practical, which, for temperature differentials of 20°C, is equivalent to ±0.5% accuracy. Therefore, using calorimetric measurement techniques, the output power of AM transmitters can be measured with total accuracies better than ±2%. The RF output amplifier efficiency factor F, referred to in FCC Rules Part 73.51(e), can then be determined for future operating and proof of performance reference. However, even with this method of determining the factor F, an accuracy of less than ±2% cannot be maintained over the life of the equipment. The required FCC transmitter voltage and current meters have a ±2% accuracy, in addition to the multiplied power accuracy of approximately ±4%. When accuracy levels are combined, this yields a total uncertainty of ±6% for the factor F. Still, this is considerably better than the accuracies obtainable by the direct measurement technique.

Operating Efficiency and Input Power

The determination of transmitter input power should be done under actual or simulated operating conditions. This requires program or simulated program signals for the transmitter modulating source during the period of power input measurement. The preferred method of measuring ac input power is by use of a standard rotating disk watt-hour meter. The measurement accuracy of these familiar watt-hour meters is typically better than 0.5%, better than four times the accuracy of any other conventional direct ac power measurement technique, and can be obtained with accuracies better than 0.1%. The watt-hour metering system should be connected in the main power feed line to the transmitter under test. Sinusoidal test signals, useful for other tests of transmitter performance, are not recommended for power consumption tests because of the distinctly opposite statistical characteristics of sinusoidal signals and typical voice and music program material and the effect this difference has on actual operating power consumption measurements. This difference between periodic sinusoidal signals and mathematically random types of signals such as human voice and/or music has been known for decades. The effect this difference produces in AM broadcast transmitter power consumption and operating efficiency measurements, however, was first documented in 1980, by investigators in Europe and was later verified and further explained by investigators in the US and other countries.[15,16]

The critical difference between sinusoidal signals and typical program types of signals is explained in Figure 4.1-18 which shows the amplitude density characteristics of a sinusoidal waveform (U-shaped curve) and an amplitude density characteristic of typical broadcast program modulation (pyramid curve). The totally different shapes of the two curves in Figure 4.1-18 cause the measured transmitter efficiency at identical RMS modulation levels also to be quite different. The pyramid shaped curve in Figure 4.1-16 was generated in 1984 from measurements taken off the air of five differently formatted FM radio stations in Dallas, TX. Data was taken for 24 hours continuous for each station and averaged for presentation as shown in Figure 4.1-18. FM stations were used in the collection of the data because of their consistent day/night signal levels, symmetrical modulation characteristics

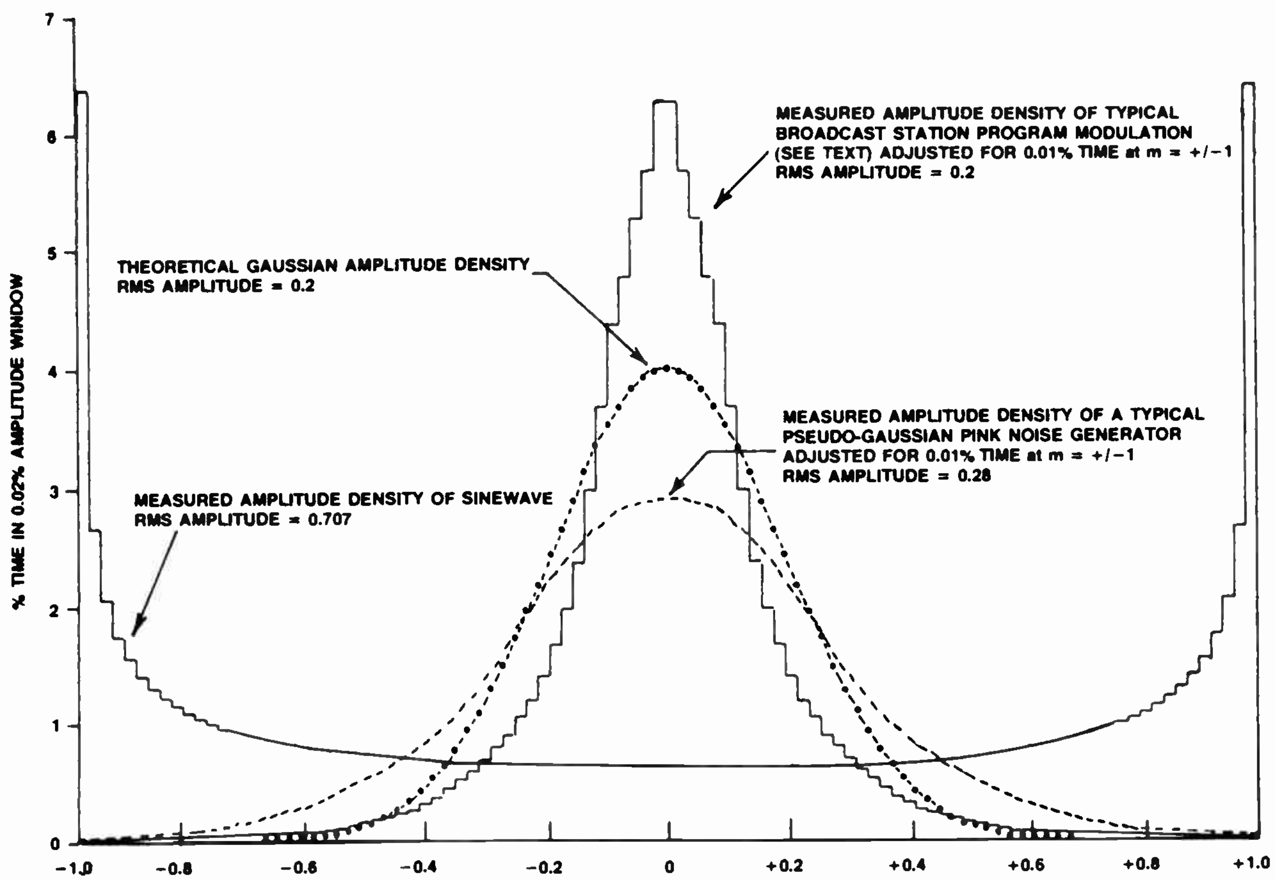


Figure 4.1-18. Amplitude density of sinusoidal test signals and typical program or noise test signals.

for easier comparison to the sine wave, and omnidirectional emission pattern. Some transmitters, or more correctly modulation systems, are more efficient with program modulation than with equivalent RMS sinusoidal test signals while others have poorer efficiency with program modulation. It has been effectively shown that modulation systems or techniques that yield higher carrier efficiencies also yield higher program efficiencies (lower program power consumption than with an equivalent RMS sinusoidal modulation test signal).[16] It is obvious that the program input power consumption test is the one most meaningful to broadcasters whose purpose it is to broadcast programs, rather than sinusoidal tones. In this discussion, the terms *program efficiency* and *program input power consumption* have generally been given equal weighted value. It is correct to equate the weighted value of these two parameters for transmitter to transmitter comparisons if it is assumed or defined that average transmitter-to-transmitter output power is constant for a defined program input. Such an assumption is correct except that transmitters with excessive transient overshoot will have correspondingly less average modulated RF power output for given peak levels of processed program modulation. Experimental methods have been devised to measure average long term transmitter output power (RF kilowatt-hour output).[17] With such equipment it would be convenient to accurately compute actual program efficiency knowing both kilowatt-hours input and kilowatt-hours output. Until such equipment becomes commercially available, however, only the transmitter energy input can be accurately measured with single-phase and multi-phase kilowatt-hour instruments.

Figure 4.1-18 also shows a third curve (dashed curve) which is the amplitude density function of a popular pink noise generator. It is recommended that either a pink noise signal with amplitude density characteristics similar to the curve in Figure 4.1-16 or a recording of actual program modulation be used as a program source for factory input power consumption tests. Program processing equipment, if used, and peak modulation levels should be adjusted to satisfy normal station operating procedures and transmitter input power measurements then taken during a 30 minute (minimum) segment of the program source material. (Refer to FCC Rules Section 73.1570(a) and (b) for modulation setup procedures.) The average input power determined by the method described will be very close to the transmitter power consumption during its total operating life and therefore can be used to accurately predict actual operating energy costs. This is the only method of transmitter power consumption tests that will provide accurate energy consumption forecasting data.

IMPORTANT AM STEREO TRANSMITTER CHARACTERISTICS

AM stereo utilizes a form of phase modulation for encoding the stereo signal onto the carrier of the AM signal. It is for this simple reason that the most important transmitter characteristic affecting AM stereo operation is IPM. Excessive IPM can affect stereo separation, single channel distortion and the occupied bandwidth of a stereo transmission—with the most significant of the three being single channel distortion.

Recognizing the various causes of IPM is the first step in correcting the trouble. There are many potential sources of IPM but the most common source is incorrect amplifier neutralization, either of a final modulated RF amplifier or of a lower power driver stage. The solution is, of course, to perform better neutralization of the offending amplifier stages. This is easier said than done in most cases. Since the inception of AM stereo broadcasting, manufacturers of the transmitters have paid more attention to the problems of IPM and in most cases have reduced levels of IPM in current production model transmitters to acceptable levels for AM stereo operation. In some cases, for especially difficult problems, field engineers from the manufacturers factory must make special on-site adjustment of individual transmitters which are particularly susceptible to excessive IPM. Several engineering consulting firms have developed special knowledge in the AM stereo field, having collected data on many transmitters in current use some of which have been out of production for several years. It is suggested that such a firm be contacted to solve particularly difficult neutralization procedures on old or new equipment which may not have been satisfactorily performed by the original manufacturer.

Phase Noise

Residual phase modulated noise cannot normally be detected by a standard AM broadcast receiver employing envelope detection, while stereophonic AM receivers are sensitive to PM noise. PM to AM conversion can occur on the medium wave band over multi-ionospheric *hops* (nighttime sky wave propagation) which can then be detected by receivers employing standard envelope detection. Very early transmitters sometimes produced more residual phase noise sidebands than those produced by the desired program AM. Nighttime reception of distant stations sometimes was accompanied by a 120–360 Hz roar which was caused by phase modulation from the master oscillator filament supply. A somewhat subdued roar can be heard even today from many shortwave stations using older transmitting equipment, especially on the higher shortwave bands where multihop propagation is more prevalent.

Significant phase modulated noise in modern transmitters is virtually nonexistent due to the use of high quality quartz crystal oscillator and synthesizer circuitry. Phase noise modulation of 0.6 degrees (0.01 rad) average is fully acceptable for monophonic AM broadcasting. Phase noise modulation of approximately 0.2° (0.0035 rad) average is usually considered acceptable for AM stereophonic broadcasting. As with measurements of IPM, an HP-8901, or equivalent, modulation analyzer is recommended for phase noise modulation measurements.

Stereophonic Phase/Gain Equalization

Standard production exciters for AM stereo systems may incorporate built-in circuitry designed to approximately match the phase and gain characteristics of the normal monophonic transmitter transmission path to the transmission path for the encoded stereo signal. Transmitters which have excessive in band non-linear phase characteristics sometimes require special outboarded phase/gain equalization networks to achieve optimum stereo performance.

International Shortwave Broadcast Transmitters

International shortwave broadcasting in the United States and North America began at about the same time as medium wave broadcasting, about 1920, but did not have any substantial growth until pre-World War II propaganda activities by the influential nations. Until this growth in the mid-1930's, shortwave broadcasting was done primarily by amateur and special industry groups using the higher frequency bands for experimentation and hobby interests. Almost immediately after the invasion of Pearl Harbor, accelerated shortwave broadcast activity in the United States began by the Office of War Information (OWI) which later assumed the wartime responsibility of the then inexperienced Voice of America in June 1942. Later, in mid-1943 the Armed Forces Radio Service began shortwave relay broadcasts to France and other parts of Europe from bases in England. After WWII, the VOA resumed control of their wartime expanded facilities and, together with virtually all nations, began continuous operations of native and foreign language broadcasts for information exchange and general entertainment purposes. Other private interest in shortwave broadcasting also began to expand after the war, primarily from various religious and politically based organizations seeking means to advance their cause or provide information and service not otherwise available by government operations. Commercial shortwave broadcasting in the United States has generally not been profitable primarily because of the limited audiences in North America given the popularity of AM, FM, and TV broadcasting. The quality of signal in shortwave listening is not of the same quality as in either of the three other more popular modes of broadcasting, due to the long distances covered and atmosphere-generated fading and noise. Therefore, shortwave broadcasting in North America attracts a different kind of listener, one who is more dedicated to the hobby of shortwave listening and therefore more willing to tolerate the disturbances associated with shortwave listening. Shortwave listening in the world outside North America is very popular, and necessary, among large segments of the population as a source of news, sociopolitical or religious viewpoint and entertainment.

Shortwave Transmitters

Shortwave broadcast transmitters are similar in many respects to medium wave transmitters and very different in others. The similarities are in methods of modulation, control, and monitoring. The differences are, generally, that shortwave transmitters are higher power, more complex in tuning, and more difficult to operate and maintain than medium wave standard broadcast transmitters. Although there are numerous exceptions, the general rule, is that the minimum usable carrier power level is 50 kW and the maximum economical carrier power level from a single transmitter is 250 kW to 500 kW.

It is not unusual for a shortwave transmitter to operate on five to ten separate frequencies every broadcast day. Modern broadcasting schedules of the prestigious broadcasting organizations are very tight, thus necessitating built-in automatic frequency changing circuitry which allows the transmitter to tune to several programmed frequencies in approximately 10 to 30 seconds, with minimum or no operator intervention. The trend in shortwave transmitter operation is toward unattended or minimum attended sites, with program and frequency changes done by remote and/or computer control.

Single Sideband and Digital Broadcasting on the International Shortwave Bands

Since about 1964, ITU has encouraged the adoption of a form of single sideband (SSB) as the standard modulation system for international shortwave broadcasting. This recommendation has been prompted due to the ever increasing congestion in the shortwave broadcast bands, the competitive increases in transmitter power levels, which contributes to the congestion and the realization of the increased efficiency of the SSB mode of transmission. A World Administrative Radio Conference Committee, meeting in Geneva in January and February 1984, adopted a specific and detailed 20-year plan for conversion to SSB on the international shortwave broadcast bands. The committee report addressed necessary changes in both transmitter and receiver technologies, thus creating the beginning and emphasizing the importance of the changes which can double the available channel space and improve reception quality in the current international broadcast bands.

Interest in compatible and discrete digital broadcasting on the international shortwave bands has been the natural outgrowth of similar interests in digital medium wave AM and VHF-FM broadcasting. The advantages of shortwave for cross-border broadcasting is well known. Since the end of the Cold War in the early 1990s, the popularity of shortwave broadcasting has declined in formerly closed societies, as reliable sources of independent news and thought have become more competitive in those countries. Nevertheless, almost every national administration has a desire to maintain short and long distance cross-border broadcasting capability in order to maintain stability in world politics. The popularity of shortwave dramatically increases whenever global or regional political instabilities develop, and digital shortwave broadcasting holds the promise to convert this medium to a higher technical quality. Various national administrations and private industry groups have begun independent as well as cooperative efforts to explore and develop the technology required for digital audio transmission and reception on the shortwave bands. Development of digital shortwave technologies could completely preempt the development of single sideband broadcasting.

AUDIO PROCESSING AND PREEMPHASIS

Audio processing technology progressed significantly in the 1980s. Manufacturing and research companies developed sophisticated circuitry to further enhance the state-of-the-art in speech and program audio processing toward the goal of producing maximum perceived loudness and quality by the listening public. It is not the intention here to discuss in detail the various processing schemes and advantages that each may offer. However, it is important that a discussion of AM transmitter technology also make mention of AM receivers, and the effect of audio processing on them.

Another important achievement in the 1980s was that the transmitter and audio processor equipment manufacturers, broadcasters, and receiver manufacturers came together in a forum group called the National Radio Systems Committee (NRSC), now jointly sponsored by the National Association of Broadcasters (NAB) and the Consumer Electronics Manufacturers Association (CEMA). Out of this forum came the definition and formal recognition of certain system inconsistencies that were a detriment to the growth of AM broadcasting. As a result of the NRSC's work, new and very important standards involving both the transmission and receiving ends of the system for AM broadcasting were developed.

The major system inconsistency that exists in AM broadcasting had been known for some time by many individuals in the industry. The offending system parameter could be called the transmitter/receiver interface. Because of the nature of nighttime skywave propagation on the standard medium wave AM band, and the adjacent and sometimes next adjacent and even not too uncommon next-next adjacent channel interference created by such propagation, receiver manufacturers had for many years severely restricted the intermediate frequency (i-f) and audio bandwidth of the receivers it produced for the AM band. In some cases the bandwidth restriction was severe (−20 to −30 dB frequency response at 5 kHz was not unusual). Yet, the broadcasters continued to broadcast wideband audio out to at least 10 and often 15 kHz in an effort to provide the public with the highest quality signal possible. The NRSC addressed the transmitter/receiver interface mismatch and produced a recommended standard of restricted transmitted bandwidth to 10 kHz which would encourage receiver manufacturers to widen the i-f and audio frequency response of the receivers correspondingly. The solution was the only one that could be made in light of the present usage and congestion in North America on the medium wave AM band. The standard that the NRSC proposed was a voluntary standard only and required the voluntary compliance of a majority of North American broadcasters and manufacturers of AM receiver products for the North American market. This voluntary compliance was enthusiastically instituted by more than 1,000 private AM broadcasters in the U.S. and Canada, prompting the FCC to make the NRSC standard a requirement by all U.S. AM broadcast stations. The FCC requirement took effect on June 30, 1990. Part of the standard generated by the NRSC involves the recommendation of an AM pre-emphasis curve to be used in the audio processing equipment at the transmitter end of the link. The proposed preemphasis curve is called a *Modified 75 Micro-second AM Preemphasis Curve* and exhibits approximately 1 dB boost at 1 kHz and approximately 10 dB boost at 10 kHz, but followed or accompanied by a single pole low pass filter with a break frequency of 8.7 kHz to reduce the peak boost at high frequencies (see Figure 4.1-19). The proposed matching deemphasis curve in the receiver can be achieved in the i-f or audio stages of the receiver, or a suitable combination of the two (see Figure 4.1-20).

The recommendation by the NRSC and following supportive action by the FCC is at least an important first step in achieving maximum system bandwidth on the North American AM broadcast band. The issue and subject of high audio frequency preemphasis is not without some controversy. Some broadcast engineers believe the standard is good in concept but does not go far enough, believing that high frequency boost should be limited somewhat lower than 10 kHz to further reduce or possibly eliminate adjacent channel interference. The European Broadcasting Union (EBU) in conjunction with the ITU-R in Geneva, Switzerland recommends moderate high frequency preemphasis to 4.5 kHz (European AM channels are spaced at 9 kHz intervals), followed by a sharp cutoff low pass filter

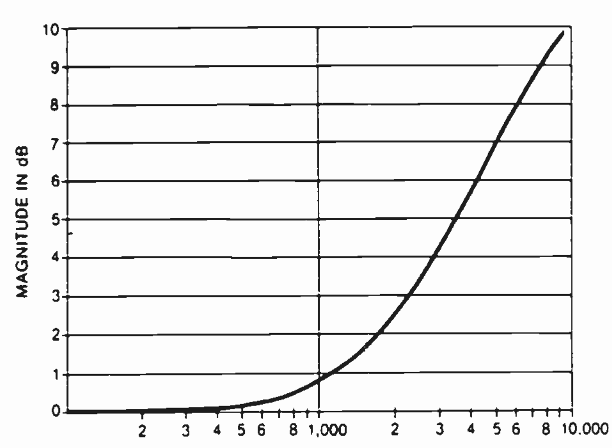


Figure 4.1-19. Modified 75 micro-second AM standard preemphasis curve.

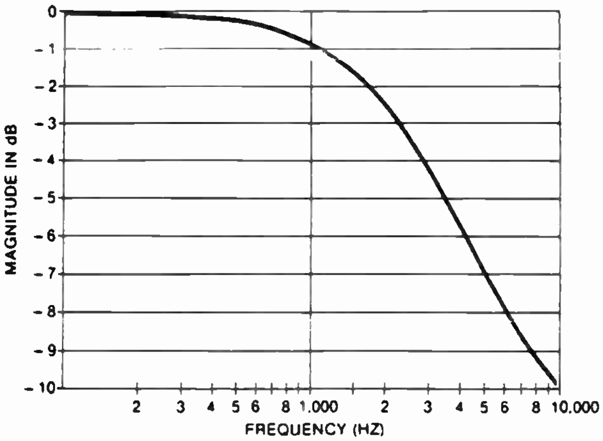


Figure 4.1-20. Modified 75 micro-second AM standard deemphasis curve.

at frequencies above 4.5 kHz. An identical standard of 4.5 kHz maximum sideband width is also recommended by the ITU-R for shortwave broadcasting where the channel spacing is 5 kHz.

The NRSC also generated a second standard pertaining to measurement of actual occupied bandwidth of a licensed AM broadcast station. The FCC ordered that all U.S. AM broadcast stations comply with this second standard, NRSC-2, by June 30, 1994. The full text of NRSC standards 1 and 2 follow as Appendices A and B.

REFERENCES

1. Carson, John R., "Notes on Modulation Theory," *Proceedings of the Institute of Radio Engineers,* Volume 10, No. 1, pp. 57, June 1925.
2. Heising, J. A., "Production of Single Sideband for Trans-Atlantic Radio Telephone," *Proceedings of the Institute of Radio Engineers,* Vol. 13, No. 3, pp. 291, June 1925.
3. Inaugural Address of Dr. Lee DeForest upon becoming President of the Institute of Radio Engineers, *Proceedings of the Institute of Radio Engineers,* No. 7, pp. 1121, July, 1930.
4. Carson, John R., "The Equivalent Circuit of the Vacuum Tube Modulator," *Proceedings of the Institute of Radio Engineers,* Vol. 9, No. 4, pp. 243, August 1921.
5. Heising, R. A., "Modulation in Radio Telephony," *Proceedings of the Institute of Radio Engineers,* Vol. 9, No. 3, pp. 305, June 1921.
6. Chireix, Henry, "High Power Outphasing Modulation," *Proceedings of the Institute of Radio Engineers,* Vol. 23, No. 11, pp. 1370, November 1935.
7. Doherty, W. H., "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the Institute of Radio Engineers,* Vol. 24 No. 9, pp. 1163, Sept. 1936.
8. Weldon, J. O., "Amplifiers," U.S. Patent No. 2,836,665, May 1958.
9. Heising, R. A., "Transmission System," U.S. Patent No. 1,655,543, January 1928.
10. Kretzmer, Ernest R., "Distortion in Pulse Duration Modulation," *Proceedings of the Institute of Radio Engineers,* Vol. 35, No. 11, pp. 1230, November 1947.
11. Terman, F. E. and Woodyard, John R., "A High Efficiency Grid-Modulated Amplifier," *Proceedings of the Institute of Radio Engineers,* Vol. 26, No. 8, p. 929, Aug. 1938.
12. Sainton, J. B., "High Efficiency Amplifier and Push-Pull Modulator," U.S. Patent No. 3,314,024, April, 1967.
13. Panter, Philip F., *Modulation, Noise, and Spectral Analysis,* McGraw-Hill, Chapter 20, pp. 640–641, 1965.
14. Doherty, W. H., "Operation of AM Broadcast Transmitters into Sharply Tuned Antenna Systems," *Proceedings of the Institute of Radio Engineers,* Vol. 37, No. 7, pp. 729, July 1949.
15. Sempert, M. and Tschol, W., "Efficiency of High Power Broadcasting Transmitters in Regular Programme Service," *BBC Review,* July 1980.
16. Woodard, G. W., "Efficiency Comparison of AM Broadcast Transmitters," *IREE Journal* (Australia), Vol. 2, No. 2, June 1982.
17. Woodard, G. W., "Simulating Typical Program Modulation Measurements of Operating Efficiency and Modulation Capability of AM Broadcast Transmitters," *Radio and Electronic Engineer (UK),* Vol. 53, No. 9, pp. 325.

NATIONAL RADIO SYSTEMS COMMITTEE

Electronic Industries Association

National Association of Broadcasters

INTERIM VOLUNTARY NATIONAL STANDARD

1. 75 µS AM Broadcast Transmission Preemphasis
2. Complimentary 75 µS AM Receiver Deemphasis
3. 10 kHz AM Transmission Bandwidth
4. Five-year Review Provision

January 10, 1987

Sponsored by the Electronic Industries Association and the National Association of Broadcasters

NRSC Co-convenors:

Charles T. Morgan
Susquehanna Radio Corp.
York, Pennsylvania

Bart Locanthi
Pioneer North America, Inc.
Duarte, California

Alan Boyer
Sony Corp. of America
Kansas City, Missouri

NRSC Subgroup Co-convenors:

John Marino
NewCity Communications
Bridgeport, Connecticut

William F. Gilbert
Delco Electronics Corp.
Kokomo, Indiana

TABLE OF CONTENTS

§ 1. SCOPE

The National Radio Systems Committee (NRSC) is a joint Committee composed of all interested parties including representatives of AM broadcast stations, AM receiver manufacturers, and broadcast equipment manufacturers. This document describes an interim[1] voluntary national standard that specifies the preemphasis of AM broadcasts, the deemphasis of AM receivers, and the audio bandwidth of AM stations prior to modulation. The standard applies to AM monophonic and AM stereo L+R transmissions, and to dual bandwidth and single bandwidth AM receivers. Compliance with the standard is strictly voluntary. To the NRSC's knowledge, no industry group or entity is or will be adversely affected by issuance of this document. Every effort has been made to inform and accommodate any and all interested parties. The NRSC believes that implementation of the standard will reduce AM interference, increase useful AM service areas, and encourage the production of higher fidelity AM receivers.

A five year review provision is established.

§ 2. INTRODUCTION

On September 5, 1985, the NRSC adopted a resolution to study proposals to standardize AM transmission preemphasis and AM receiver deemphasis with the objective of establishing an industry-wide AM preemphasis/deemphasis voluntary standard. After twelve months of study, on September 10, 1986 the NRSC released a draft voluntary standard that proposed a specific AM preemphasis/deemphasis curve as well as a 10 kHz standard AM bandwidth. The bandwidth specification evolved from NRSC deliberation on the causes and cures of AM interference, and ways to technically encourage the production of higher fidelity AM receivers. After a three month public comment period, the NRSC, on January 10, 1987, formally adopted this standard and authorized its publication by the National Association of Broadcasters and the Electronics Industries Association.

The purpose of the NRSC voluntary standard is to create a transmission/reception system where (1) AM broadcast stations will know, with certainty, the likely audio response characteristics of AM receivers, and (2) AM receiver manufacturers will know, with certainty, the likely audio response characteristics of AM broadcasts. A "matching" of preemphasis and deemphasis is expected to improve the consumer's overall satisfaction with the technical quality of listening to AM radio. The NRSC believes that the public interest is served by establishing a compatible transmission/reception system and the accompanying improvement of AM broadcasts and AM receivers.

This document also describes a specification for the maximum audio bandwidth transmitted by AM broadcast stations. Implementation of a bandwidth specification will reduce second-adjacent channel interference and thereby lead to (1) a significant reduction of second-adjacent channel interference as perceived on "wideband" AM receivers; (2) a corresponding increase in the interference-free service areas of AM stations; and (3) an incentive for the further building of dual bandwidth AM "wideband" receivers.[2] Analysis by a subgroup of the NRSC has shown that there would be little if any detrimental effect on today's "narrowband" AM receivers upon the implementation of this voluntary standard.

§ 3. BASIC DEFINITIONS

A. Preemphasis. The boosting of high audio frequencies prior to modulation and transmission.

B. Deemphasis. The attenuation of high audio frequencies during the process of reception and demodulation.

C. "Narrow" receivers. A subjective term to describe receivers with typical combined RF, IF and AF response characteristics of -10 dB at 4.2 kHz, -20 dB at 6.0 kHz. Response characteristics of narrow AM receivers are known to vary widely.

D. "Wideband" receivers. A subjective term to describe receivers with typical combined RF, IF, and AF response of -6 dB

---

1. The standard is described as "interim" until it is submitted and approved by the American National Standards Institute.

2. First Adjacent channel interference considerations may continue to discourage the building of single bandwidth "wideband" receivers; however, the extent and nature of this form of interference has not been fully studied by the NRSC.

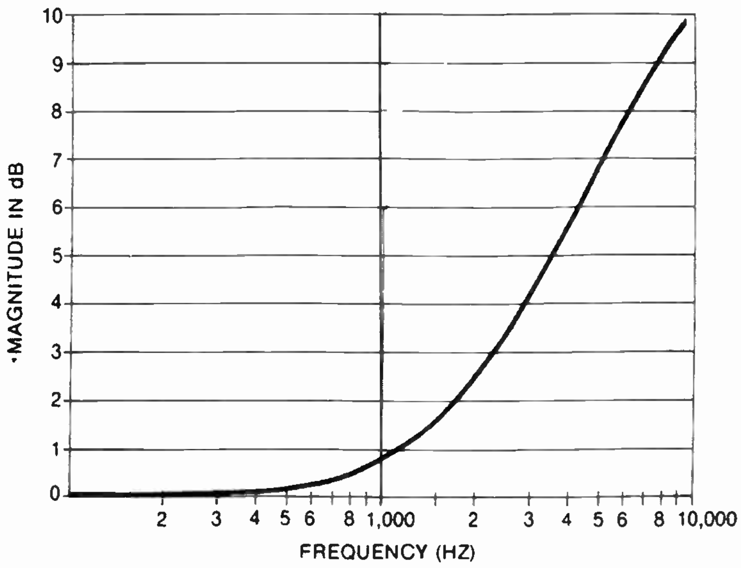


Figure 1. Modified 75μs AM Standard Preemphasis Curve

Technical Information

Frequency	Magnitude (dB)	Phase (deg)	Group Delay (sec)	Frequency	Magnitude (dB)	Phase (deg	Group Delay (sec)
50	0.00	1.0	-5.6669E-005	5000	6.92	37.1	2.3048E-006
100	0.01	2.0	-5.6547E-005	5500	7.41	36.6	3.3525E-006
400	0.14	8.0	-5.4175E-005	6000	7.85	35.9	4.0592E-006
700	0.42	13.7	-4.9467E-005	6500	8.24	35.2	4.5169E-006
1000	0.81	18.7	-4.3318E-005	7000	8.58	34.3	4.7926E-006
1500	1.63	25.5	-3.2247E-005	7500	8.89	33.4	4.9357E-006
2000	2.54	30.4	-2.2343E-005	8000	9.16	32.5	4.9823E-006
2500	3.44	33.6	-1.4509E-005	8500	9.41	31.6	4.9595E-006
3000	4.28	35.7	-8.6612E-006	9000	9.62	30.8	4.8871E-006
3500	5.05	36.9	-4.4133E-006	9500	9.82	29.9	4.7801E-006
4000	5.75	37.4	-1.3702E-006	10000	10.00	29.0	4.6495E-006
4500	6.37	37.4	7.8900E-007				

at 6 kHz, -10 dB at 8 kHz. Response characteristics of wide AM receivers are known to vary widely.

E. "Excessive" Preemphasis. Preemphasis that produces no discernable benefit when received by a "narrow" receiver but increases interference to adjacent channel AM stations.

§ 4. AM TRANSMISSION PREEMPHASIS

§ 4.1. In General. AM preemphasis is the boosting of high audio frequencies prior to modulation and transmission. Today, most AM stations use preemphasis to varying extents. This preemphasis is employed in an attempt to compensate for the "narrow" response of most AM receivers. If AM preemphasis is not controlled, one station may interfere with AM receivers listening to neighboring stations located on adjacent AM channels. Whether such interference is objectionable will depend on (1) the response characteristics of the AM receiver, (2) the amount and nature of transmission preemphasis, (3) the extent to which the AM station is employing compression/ limiting techniques, and (4) whether the AM transmission system is bandlimited in the audio processor, transmitter or antenna.

Preemphasis is useful for improvement of the AM transmission-reception system audio response only to a limited extent for receivers using IF transformers. Many receivers using ceramic filters with narrow response characteristics can not be improved by use of excessive preemphasis. These receivers can not "hear" the transmission of preemphasized high audio frequencies. But excessive preemphasis will foster adjacent channel interference and cause wideband radios to sound shrill or strident.

§ 4.2. Description of the Modified 75 uS Preemphasis Curve. Each AM broadcast station shall broadcast with audio preemphasis as close as possible (within the capabilities of the station's transmission system) to the recommended standard, without exceeding it. The curve applies for audio frequencies up to 10 kHz.

The NRSC proposed standard AM transmission preemphasis curve is shown in Figure 1. The curve describes the recommended net transmission system static audio response of an AM station.

The recommended preemphasis curve is a single zero curve with a break frequency at 2122 Hz. It is similar to the 75 uS curve used for FM broadcasting. To reduce the peak boost at high frequencies, a single pole with a break frequency of 8700 Hz is employed. NRSC analysis has shown that the proposed curve is compatible with most existing AM receivers.

§ 4.3. Methods for Determining Performance. The NRSC AM preemphasis curve is a static curve, and cannot be measured dynamically. NRSC studies have shown that the dynamic and non-linear functions performed by most AM station audio processors will modify any given preemphasis curve. In addition, it is the audio response of the entire AM transmission system that indicates performance in accordance with the standard. For these reasons, measuring a station's preemphasis curve for the purpose of determining compliance with the NRSC standard shall be performed in accord with the following specifications:

§ 4.3.1. Use of Audio Tones. Compliance with the curve shall be measured by sweeping the station's transmission system with audio tones. The dynamic functions of the AM station's processor, but not the frequency shaping circuits, must be disabled (i.e., in "proof" mode).

§ 4.3.2. Location of Measurement. The net transmission system audio response is best measured by detecting the over-the-air signal. This will ensure that the AM transmitter and antenna combination is faithfully reproducing the preemphasized audio.[3] Alternatively, if the transmitter and antenna combination is reasonably broadband, performance can be determined by static measurement of the audio signal prior to modulation.

§ 5. AM RECEIVER DEEMPHASIS

§ 5.1. In General. Receiver deemphasis results from the selectivity

3.However, the deemphasis characteristics of the device used to demodulate the AM transmission must be accounted for. Additionally, some AM stations with transmitter or antenna problems may not be able to pass preemphasized audio without introducing "splatter" interference and/or overmodulation. If a particular AM station transmission system cannot "handle" the NRSC recommended curve, it is suggested that a lower amount of preemphasis be used until the system problems are corrected to allow the NRSC curve to be faithfully implemented.

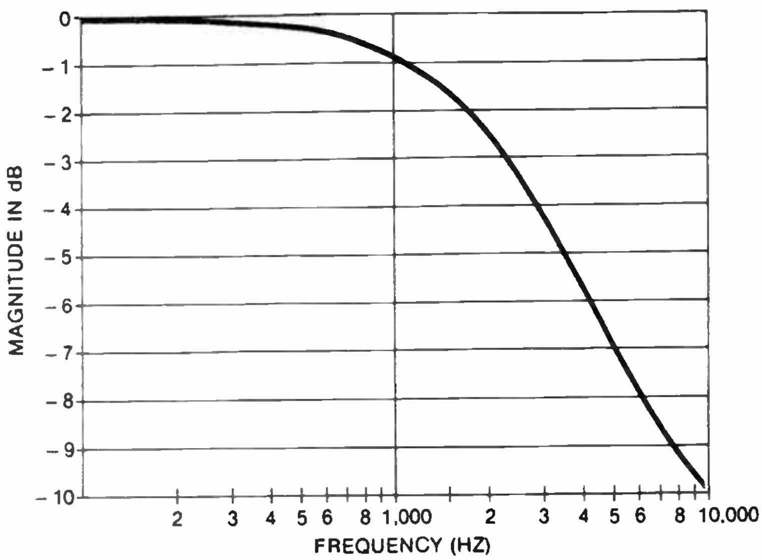


Figure 2. Modified 75µs AM Standard Deemphasis Curve

characteristics of a receiver's RF and IF stages and the response characteristics of the receiver AF section. A standard deemphasis curve permits AM stations to know, with certainty, the likely overall response characteristics of AM receivers.

§ 5.2. Description of the Standard Deemphasis Curve. AM receivers shall complement the recommended transmission preemphasis characteristic described in § 4 by incorporating a net receiver system audio response described in Figure 2. (The net system audio response of an AM receiver is the combined RF, IF, and AF audio response.) The NRSC deemphasis curve is characterized by a single pole at 2122 Hz and a single zero at 8700 Hz. It is the precise complement of the preemphasis standard described in § 4. The preemphasis/deemphasis voluntary standards apply only for audio frequencies below 10 kHz; the implementation of preemphasis/deemphasis standards produces a transmission/reception system that is essentially flat to nearly 10 kHz and limited only by the AM receiver's choice of bandwidth.

§ 5.3. Methods for Determining Performance. The deemphasis characteristic shall be determined by measuring the overall frequency response in accordance with International Electrotechnical Commission ("IEC") Publication 315.3, Clause 11.2:

(1) The receiver is brought under standard measuring conditions and the reference audio-frequency output voltage is noted. The modulation frequency is then varied and the output voltage at each frequency is noted and expressed in decibels relative to the reference voltage.

The modulation depth is adjusted at each frequency in accordance with the preemphasis characteristic of AM broadcast transmission. To avoid overmodulation at some frequencies it may be necessary to use a modulation factor of less than 30% at some frequencies.

(2) If overloading of the AF section of the receiver occurs, either the volume control attenuation should be increased or the modulation factor reduced, and a corresponding factor applied to the results.

(3) The measurements may be repeated with other values of RF input signal level and frequency.

The frequency response shall be measured for both monophonic and stereophonic reception, in accordance with the definition of the particular AM stereo system. For dual bandwidth receivers, the frequency response shall be measured in both bandwidth positions.

Results may be presented graphically, with modulation frequency plotted logarithmically as abscissa and the output in decibels as ordinate.

The frequency response can be stated as follows:

Selectivity	Frequency Response
Narrowband	50 Hz to 5000 Hz +/- X dB
Wideband or Stereo	50 Hz to 10,000 Hz +/- X dB

(Where X is the maximum deviation from the recommended frequency response, and 5000 Hz and 10,000 Hz are example frequency specifications.) The deviation X shall be of as low a value as practical. If a notch filter is used while the AM receiver is under test, the stated frequency response should be modified accordingly. Suggested modifications include (1) adding an appropriate footnote to the frequency response specification; and/or (2) lowering the upper limit value to the above "wideband" audio response specification.

§ 5.4. Notch Filters. A notch filter is a very selective filter that attenuates the spectrally pure carriers of first adjacent channel AM stations. Although an optional enhancement to an AM receiver, using notch filters is recommended. If used, the notch filter should (1) have as high a "Q" as is practical, (2) adequately suppress the interfering carriers, and (3) not unduly degrade the desired bandwidth performance of the AM receiver.

§ 6. 10 KHZ BANDWIDTH FOR AM TRANSMISSION

§ 6.1 In General. Each AM broadcast station shall modulate its transmitter with an audio bandwidth described by the specification in Figure 3. Appropriate and carefully designed audio low-pass filters as the final filtering prior to modulation can be used to implement this specification. The purpose of the bandwidth specification is to remove interference by controlling the occupied RF bandwidth of AM stations.[4]

§ 6.2. Description of the Standard. The audio bandwidth transmission standard is specified in Figure 3. The audio

4.It should be noted that the operation of non-linear AM Stereo systems theoretically may produce phase modulation components outside the desired RF bandwidth. The NRSC will examine this phenomena with the goal of determining whether such components exist and, if they do exist, whether they are objectionable. For a discussion of the detrimental effect of high audio frequencies on occupied RF bandwidth, see Klein, Modulation, Overmodulation, and Occupied Bandwidth: Recommendations for the AM Broadcast Industry, Proceedings of the 1987 NAB Engineering Conference, Dallas, Texas (April, 1987).

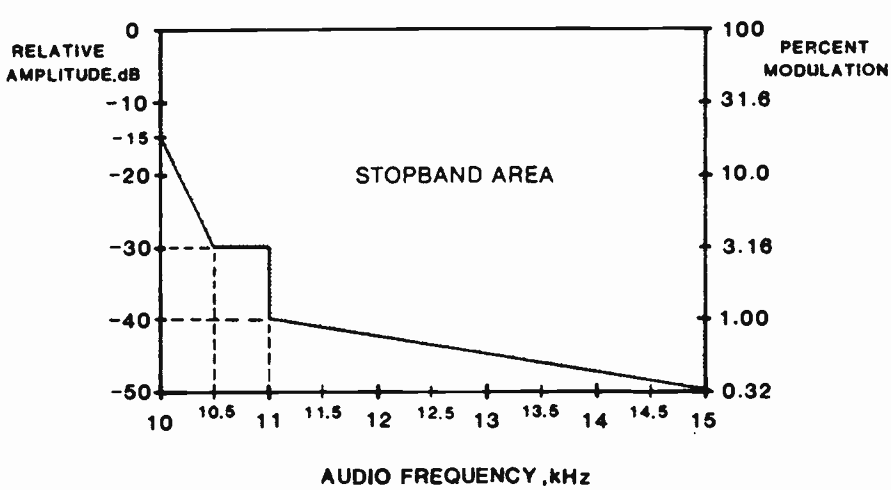


Figure 3. NRSC Stopband Specification
(Audio Envelope Input Spectrum To AM Transmitter)

envelope input spectrum to the AM transmitter shall be -15 dB at 10 kHz, smoothly decreasing to -30 dB at 10.5 kHz, then remaining at -30 dB from 10.5 kHz until 11.0 kHz. At 11.0 kHz, the audio bandwidth shall be -40 dB, smoothly decreasing to -50 dB at 15 kHz. Above 15 kHz, the audio bandwidth shall remain at least -50 dB. The reference level is 1 dB above a 200 Hz sine wave at 90% negative modulation. See Figure 3.

§ 6.3. Method for Determining Performance. An AM station is determined to be in compliance with the NRSC bandwidth characteristic by measurement of the station's audio bandwidth in accordance with the following parameters:

§ 6.3.1. Location of Measurement. Audio bandwidth measurements shall be obtained at the audio input terminals to the AM transmitter. For AM Stereo stations, audio bandwidth shall be measured at the L+R audio input terminals to the RF modulator. Note that the NRSC bandwidth standard characterizes an audio bandwidth that represents station program material that has been modified by possibly non-linear circuits in the station's audio processor. For this reason, the NRSC recommends use of a test signal that adequately characterizes typical audio program material, rather than relying on static audio test tones. However, it may still be useful to measure bandwidth statically at the time that AM preemphasis is measured.

§ 6.3.2. Use of Standard Test Signal. Audio bandwidth shall be measured using a test signal consisting of USASI (United States of America Standards Institute) noise that is pulsed by a frequency of 2.5 Hz at a duty cycle of 12.5%. See Figure 4. USASI noise is intended to simulate the long-term average spectra of typical audio program material. Pulsing of the noise is intended to simulate audio transients found in audio program material. USASI noise is a white noise source[5] (i.e. noise with equal energy at all frequencies) that is filtered by (1) a 100 Hz, 6 dB per octave high-pass network and (2) a 320 Hz, 6 dB per octave low-pass network. See Figure 4. A pulsed USASI noise generator is shown in Figures 5 and 6. Using the attenuator pad, the ratio of peak-to-average amplitude shall be 20 dB at the audio output of the pulser. The station's audio processor must be in normal operating mode.

§ 6.3.3. Use of Standard Measurement Devices. A suitable swept-frequency or FFT (Fast Fourier Transform) spectrum analyzer shall be used to measure compliance with the NRSC bandwidth specification.

(a) Spectrum Analyzer Setup. When a swept-frequency audio spectrum analyzer is used to measure compliance with the NRSC bandwidth specification, the analyzer's setup shall consist of:

a. 300 Hz resolution bandwidth.[6]
b. 2 kHz/horizontal division.
c. 10 dB/vertical division.
d. Reference: 1 dB above 200 Hz (sine wave) 90% negative modulation.
e. Display: maximum peak hold (or equivalent function).

The analyzer's operating span and sensitivity are adjusted as necessary to determine compliance.

(b) Fast Fourier Transform Analyzer. When a FFT analyzer is used to measure compliance with the NRSC bandwidth specification, the analyzer's setup shall consist of:

a. Reference: 1 dB above 200 Hz (sine wave) 90% negative modulation.
b. Window: Hanning.
c. Horiz. span: 20 kHz.
d. Dynamic range: 80 dB or available range.
e. Display: Maximum peak hold (or equivalent function).

5.Acceptable white noise sources include GenRad Models 1382 and 1390B; Bruel & Kjaer Model 1405; and National Semiconductor IC No. MM 5837N.

6.Note: if the audio bandwidth under test fails to meet the NRSC specification when a 300 Hz resolution bandwidth is employed, a narrower resolution bandwidth, such as 100 Hz, may be used to determine compliance; however, the sweep rate and the video bandwidth of the analyzer must be adjusted according to the manufacturer's instructions in order to assure accurate representation of the resolution bandwidth employed. If in doubt, check with the analyzer operating manual or the manufacturer. Further Note: the NRSC may suggest a different or more precise measurement standard as the industry gains experience. Spectrum analyzers that are capable of 300 Hz resolution bandwidths at audio frequencies include, but are not limited to, Tektronix Models 5L4N and 7L5; Hewlett-Packard Models 3580A, 3582A, 3585A, 8553A or B, 8566A or B, 8568A or B, 71100A; Marconi Models 2370, 2382; Rhode/Schwarz/Polarad Model CSA-240M; and a Techron Model TEF System 12.

§ 7. FIVE YEAR REVIEW PROVISION

It is the goal of the NRSC to increase the fidelity of the AM transmission and reception system from its present state to a quality level that approaches the quality available via FM broadcasting. Towards this end, the voluntary standards described in this document shall be in effect for five (5) years from the effective date of this standard. During the interim five year period, this voluntary standard will be reviewed at least once a year to determine whether the fidelity goals of the NRSC are being realized.

§ 8. EFFECTIVE DATES

These dates serve only as objectives. AM Broadcast stations and AM receiver manufacturers are expected to make a good faith effort to implement the NRSC standard.

A. AM Broadcast Stations. The effective date of this standard is January 10, 1987.

B. AM Receiver Manufacturers. The effective date of this standard is January 10, 1988.

§§§§

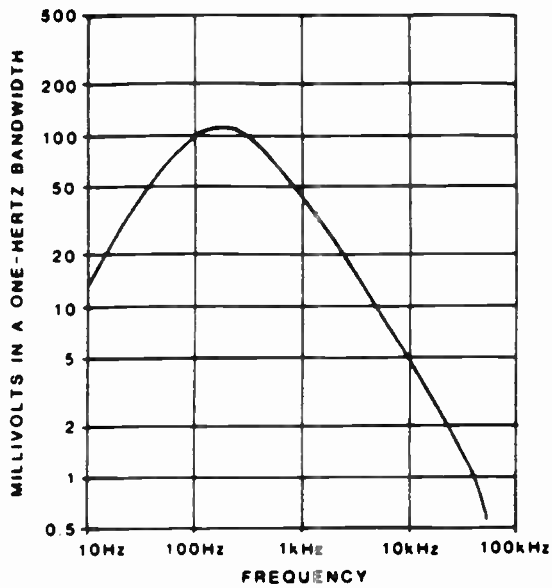


Figure 4. Spectra of USASI Noise

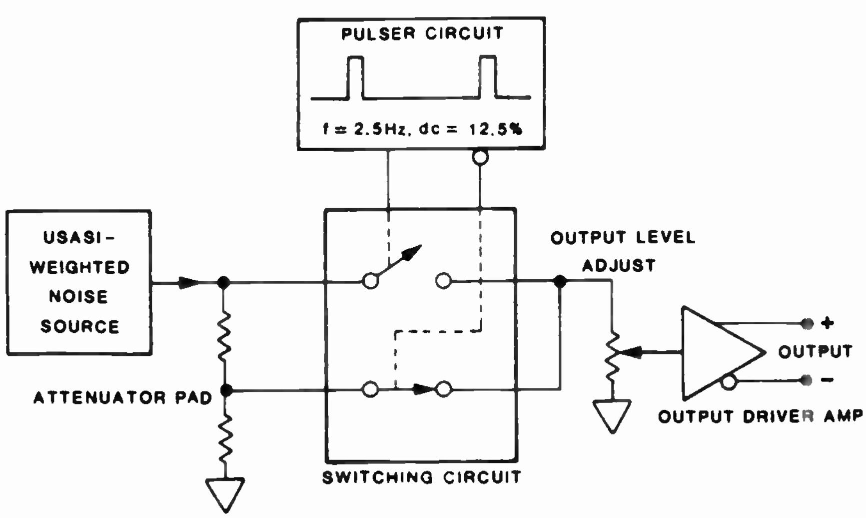


Figure 5. Pulsed-USASI Noise Generator

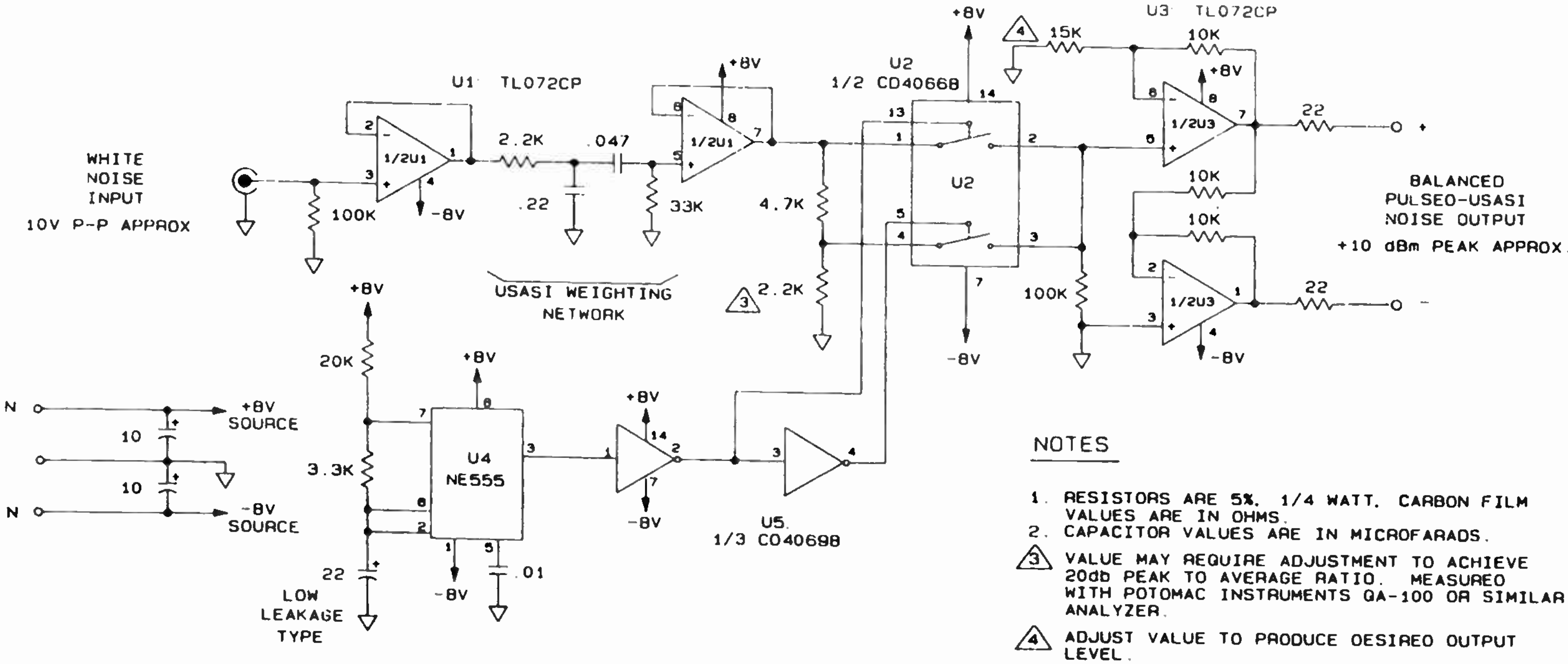


Figure 6. Application Circuit: USASI Noise Weighting/Pulser Circuit

NATIONAL RADIO SYSTEMS COMMITTEE

2001 Eye Street, N.W.
Washington, D.C. 20006
(202) 457-4975

National Association of Broadcasters
1771 N Street, N.W.
Washington, D.C. 20036
(202) 429-5346

June 1, 1988

INTERIM
VOLUNTARY NATIONAL STANDARD
EMISSION LIMITATION FOR AM BROADCAST TRANSMISSION

NRSC Chairman:

Charles T. Morgan
Susquehanna Radio Corporation
York, Pennsylvania

NRSC AM Subgroup:

John Marino
NewCity Communications
Bridgeport, Connecticut

William F. Gilbert
Delco Electronics Corporation
Kokomo, Indiana

Voluntary Standard No. NRSC-2

Sponsored by the Electronic Industries Association and the National Association of Broadcasters

TABLE OF CONTENTS

EMISSION LIMITATION FOR AM BROADCAST TRANSMISSION
(NRSC Voluntary Standard No. NRSC-2)

§ 1. Scope.

The National Radio Systems Committee (NRSC) is a joint committee composed of all interested parties including representatives of AM broadcast stations, AM receiver manufacturers, and broadcast equipment suppliers. This document describes an interim[1] voluntary national standard that specifies radio-frequency spectrum occupancy for AM broadcast stations. The standard applies to both AM monophonic and AM stereo transmissions. Compliance with the standard is strictly voluntary. To the NRSC's knowledge, no industry group or entity is or will be adversely affected by issuance of this document. Every effort has been made to inform and accommodate any and all interested parties. The NRSC believes that implementation of this standard will lead to reduced AM interference, thus providing increased service for all AM stations and an increase in quality of service to present and future AM listeners.

§ 2. Introduction.

On January 10, 1987, the NRSC authorized the National Association of Broadcasters and the Electronic Industries Association to publish an interim voluntary national standard specifying AM preemphasis, AM deemphasis and a 10 kHz AM audio bandwidth (Standard No. NRSC-1). The NRSC-1 audio standard applies to the audio signals that are intended to modulate the AM transmitter. Its purpose is to reduce second-adjacent channel interference by band limiting AM stations to a nominal

[1]The standard is described as "interim" until the test methods contained within the document can be fully verified through field tests. In addition, field test data of "splatter monitor" technology will be evaluated to determine the correlation between results obtained with such devices and the methods described within this document (see § 5.).

20 kHz occupied radio-frequency (RF) bandwidth (twice the 10 kHz audio bandwidth presented to the transmitter's modulation circuits). Implementation of the NRSC-1 audio standard alone largely achieves this purpose. However, there remain characteristics of the AM transmission process that may cause the RF occupied bandwidth to exceed a nominal 20 kHz. This document accommodates these transmission characteristics. It is in two Sections. Section 3 is a voluntary standard maximum RF occupied bandwidth of AM broadcast transmissions. Section 4 consists of a voluntary RF occupied bandwidth testing and control standard designed to insure repeatability and consistency of RF occupied bandwidth test measurements.

§ 3. RF Maximum Occupied Bandwidth Specification.

§ 3.1. Purpose. The purpose of an RF maximum occupied bandwidth specification is to control modulation products, desired or undesired, that fall outside the specified RF occupied bandwidth.

§ 3.2. Maximum Occupied Bandwidth. The maximum occupied RF bandwidth voluntary standard represents the maximum peak output of a swept-frequency spectrum analyzer IF over a minimum ten minute period.[2] The specification encompasses all spectral components caused by direct programming and all ancillary or data communications. AM broadcast stations shall occupy spectrum according to the following maximum specifications:

(see next page)

[2] It is recognized that the output of the spectrum analyzer depends on the shape of the analyzer's IF filters.

Table 1

Frequency Band Relative to Carrier (+/- kHz)	Attenuation Relative to Carrier (dB)
0 to 10	0
10 to 20	greater than -25[3]
20 to 30	greater than -35
30 to 60	greater than -(5 + 1 dB/kHz) from carrier[4]
60 to 75	-65[4]
above 75	-80[4]

(See Figures 1 and 1A (solid line) attached).

§ 3.3. Measurement Procedure.

§ 3.3.1. Use of Ordinary Program Material. Measurements of AM station spectrum occupancy shall be conducted using ordinary program material. All audio processing used in the AM station shall be in normal operating modes. The audio signal input to the AM transmitter shall conform to the NRSC audio standard adopted January 10, 1987.

§ 3.3.1.1. Use of Audio Tones. Sweeping a transmission system with audio tones is a widely accepted and respected method for gauging spectrum occupancy and for troubleshooting and adjusting AM transmission systems. The NRSC endorses audio tones for these purposes, but urges caution in the use and selection of audio tones particularly with AM stereo transmission.[5] It should be noted, however, that it is

[3] Note: the slope of occupied bandwidth in the transition region between 10 and 11 kHz is expected to parallel the NRSC-1 audio standard. Accordingly, attenuation levels in the region shall be 6 dB greater than described in the audio standard to adjust for carrier level reference.

[4] For carrier power levels between 50 and 5000 Watts, the maximum limit shall be $-(43 + 10 \log P_W)$ dB (where P_W is the carrier power in Watts) or as indicated in Table 1 and/or the attached Figure 1, whichever is lesser attenuation. For carrier power levels below 50 Watts, a -60 dBC maximum limit shall be used.

[5] The manufacturer of the particular AM stereo system employed should be consulted for the appropriate tone frequencies/modulation levels for "worst-case" condition testing.

difficult to infer the dynamic response of a transmission system while observing it in a steady-state condition.

§ 3.3.2. Use of Spectrum Analyzer. A suitable swept-frequency RF spectrum analyzer shall be used to measure compliance with the NRSC RF occupied bandwidth specification. The analyzer shall measure the over-the-air RF spectrum occupancy as perceived in the far field (i.e., at least 10 wavelengths from the antenna center). Some caution should be used in measuring spectrum occupancy with directional antennas.[6]

The analyzer's setup shall consist of:

a. 300 Hz resolution bandwidth.
b. 5, 10, or 20 kHz/horizontal division (as appropriate).
c. 10 dB/vertical division.
d. Reference: carrier peak.
e. Peak Hold: 10 minute duration minimum.
f. No Video Filter.

§ 4. RF Occupied Bandwidth Testing and Control Standard.

§ 4.1. Purpose. The NRSC recognizes that for the purposes of troubleshooting, design, and adjustment it may be desirable to use an occupied bandwidth emission standard that is coupled to a standard test signal and measurement procedure. This second measurement technique can also be utilized by transmitter manufacturers and broadcasters to provide results that may be directly correlated with each other. For these purposes, the NRSC proposes the following interim voluntary national test and control standard:

§ 4.2. Maximum Occupied Bandwidth. The RF occupied bandwidth test and control standard is a stored peak specification. With the standard test signals and measurement procedures specified below, AM stations shall occupy a maximum RF bandwidth that conforms to the following maximum specifications:

---

[6] *See* Klein, *Modulation, Overmodulation, and Occupied Bandwidth: Recommendations for the AM Broadcast Industry* (NAB, September 11, 1986) at 18-23.

Table 2

Frequency Band Relative to Carrier (+/- kHz)	Attenuation Relative to Carrier (dB)
0 to 10	0
10 to 13.5	Minimum is defined by a line with endpoints found at -25 dB/10 kHz and -35 dB/13.5 kHz[7]
13.5 to 54.5	Minimum is defined by a line with endpoints found at -35 dB/13.5 kHz and -65 dB/54.5 kHz[8]
54.5 to 75	-65[8]
above 75	-80[8]

(See Figures 1 and 1A (dotted line) attached).

§ 4.3. Measurement Procedure.

§ 4.3.1. Standard Noise Test. Measurements of AM station spectrum occupancy shall be conducted using a standard noise test signal described in the January 10, 1987 NRSC audio standard.[9] All audio processing employed in the AM station or test configuration shall be in a normal operating mode. The audio signal input to the AM transmitter shall conform to the NRSC audio standard adopted January 10, 1987.

§ 4.3.1.1. Monophonic conditions. The noise source is unmodified.

§ 4.3.1.2. Stereophonic conditions. Two independent but equivalently designed USASI-weighted noise sources are employed. Pulsing of the sources is controlled by a single control signal. The pulsed output of one noise generator is defined as L+R (mono, sum information) where the other is attenuated by 3 dB to

---

[7]See footnote 3.

[8]For carrier power levels between 50 and 5000 Watts, the maximum limit shall be $-(43 + 10 \log P_W)$ dB (where P_W is the carrier power in Watts) or as indicated in Table 2 and/or the attached Figure 1, whichever is lesser attenuation. For carrier power levels below 50 Watts, a -60 dBC maximum limit shall be used.

[9]*See National Radio Systems Committee, Interim Voluntary National Standard (NRSC-1)*, § 6.3.2 and Figures 4, 5, and 6 (January 10, 1987).

provide L-R (stereo, difference information). The signals are then matrixed to provide left and right channel information to be applied to the audio input terminals of the stereophonic audio processor employed.[10] See Figure 2 (attached).

§ 4.3.2. Use of Spectrum Analyzer. A suitable swept-frequency RF spectrum analyzer shall be used to measure compliance with the NRSC RF occupied bandwidth testing and control standard. The analyzer's setup shall consist of:

a. 300 Hz resolution bandwidth.
b. 5, 10, or 20 kHz/horizontal division (as appropriate).
c. 10 dB/vertical division.
d. Reference: carrier peak.
e. Peak Hold: 10 minute duration minimum.
f. No video filter.

§ 5. Splatter Monitor. It is understood that the NRSC does not anticipate a Spectrum Analyzer will be available for mask standard measurements in most AM radio stations. However, the current development of a low cost "splatter monitor" device may allow economical continuous monitoring of compliance with RF mask characteristics.

Such a device accommodates the following factors: (1) NRSC deemphasis, and (2) amplitude detection of in-phase and quadrature signal components found in the defined stopband region.

As these devices are evaluated and correlated with current spectrum analysis measurement techniques, a determination shall be made by the NRSC with respect to possible modification of RF mask compliance measurement methods.

§ 6. Effective Date. **June 1, 1988.**

§§§§

---

[10]The signal provided by the audio processor to the transmitter left and right audio input terminals shall not exceed single channel modulation limits as dictated by the constraints of the particular stereo system employed.

FIGURE 1

AM BROADCAST
RF EMISSION LIMITS

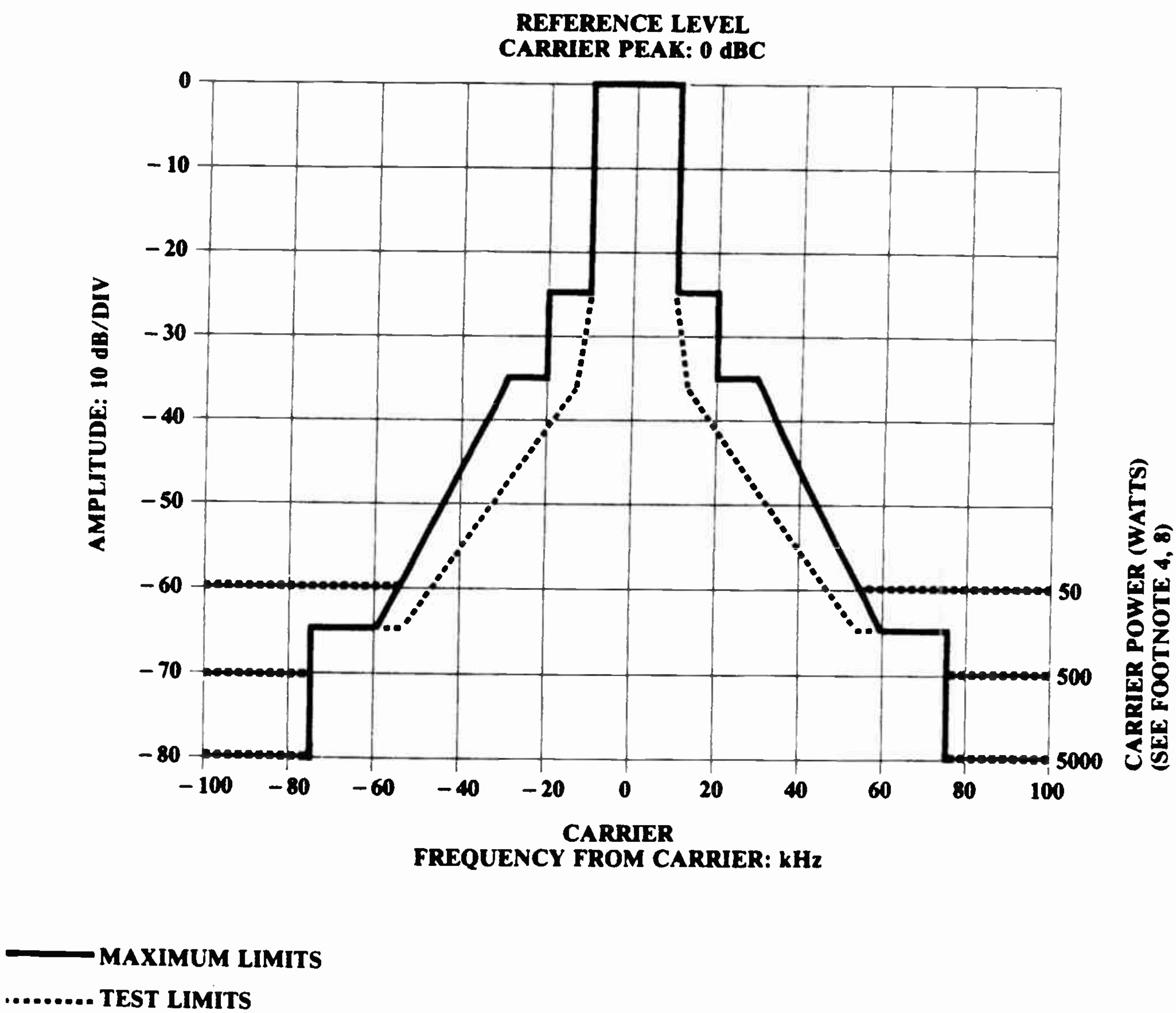
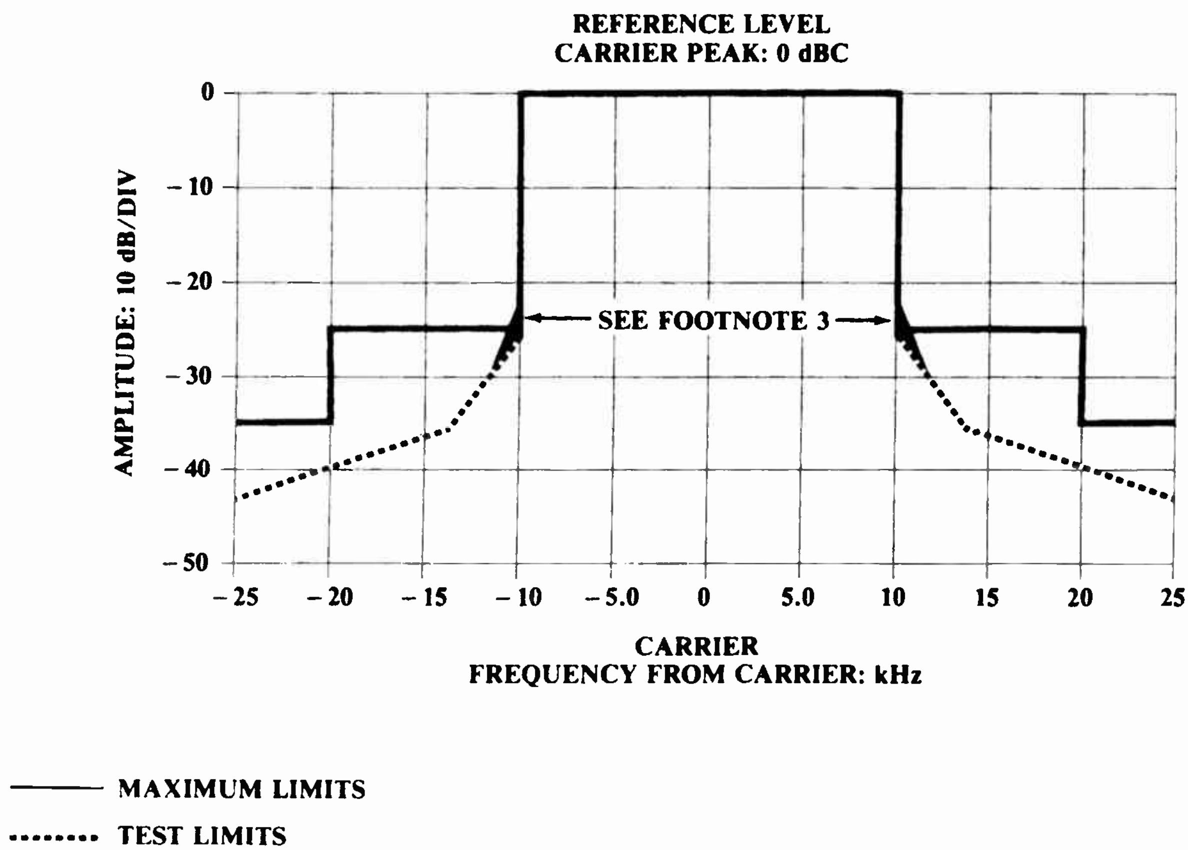
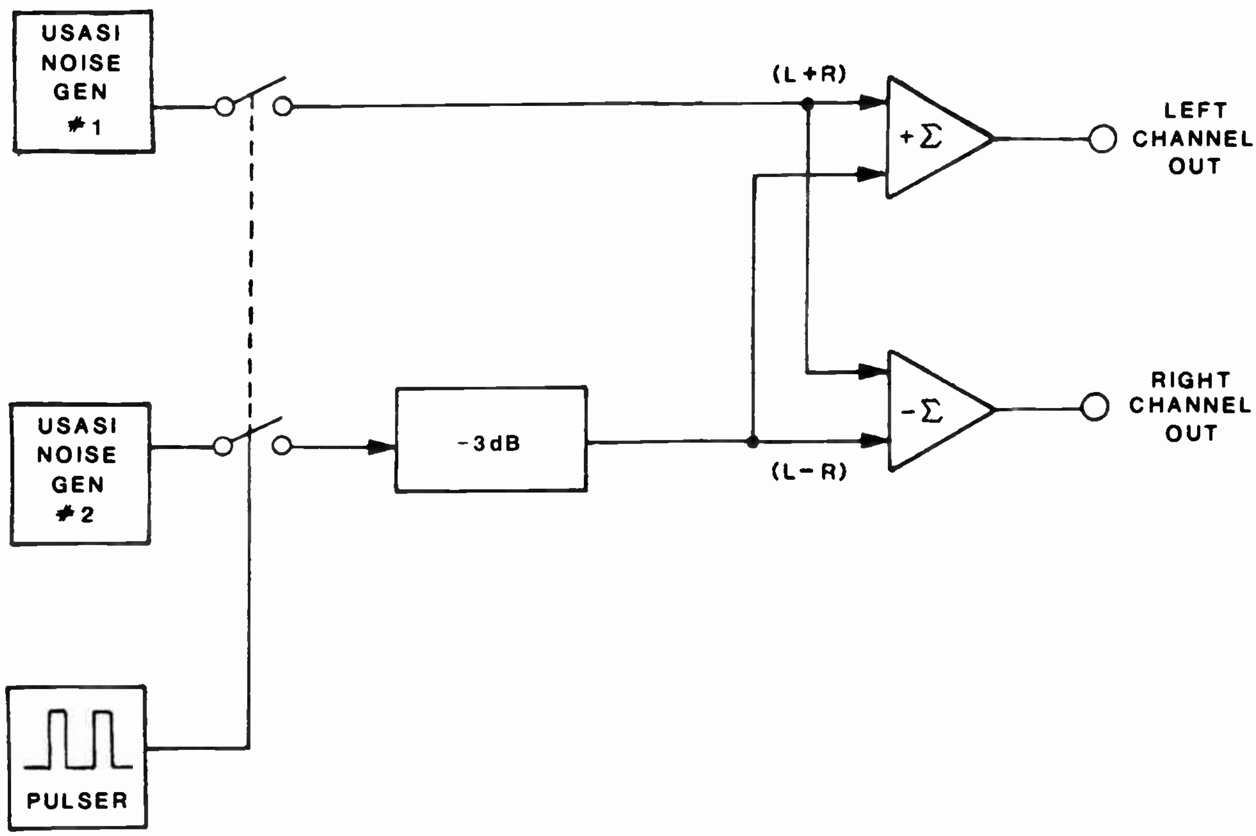


FIGURE 1 A

AM BROADCAST RF EMISSION LIMITS

(EXPANDED SCALE)




USASI NOISE GEN #1
USASI NOISE GEN #2
PULSER
-3dB
(L+R)
(L-R)
+Σ
-Σ
LEFT CHANNEL OUT
RIGHT CHANNEL OUT

4.2
AM STEREO

EDWARD J. ANTHONY
QUINCY, IL

INTRODUCTION

The acceptance and implementation of AM stereo broadcasting in the United States has been a slow and difficult transition. Beginning with the now famous FCC marketplace decision in 1981, the field of prospective AM stereo system proponents had been reduced from five (Belar, Harris, Kahn Communications, Magnavox, and Motorola) to two, namely the Kahn Independent Sideband System (ISB) and the Motorola C-QUAM System, and then to one (C-QUAM).

Recent years have shown a quiet growth in AM stereo conversion, with less controversy than was experienced in the early and mid 1980's. The broadcaster's interest turned more toward proper installation and maintenance of AM stereo than on which system to use.

The proper operation of stereo transmission for AM is relatively more difficult than its FM counterpart. This chapter will discuss the more practical aspects of AM stereo preparation, installation, and maintenance to aid the AM broadcaster in achieving a high quality stereo transmission.

STATION PREPARATION FOR AM STEREO

Depending on the particular station, part or perhaps all areas of the operation will be affected during the AM stereo installation process. It is best to consider each section of the transmission path separately, so as to make a more thorough and logical conversion.

The Audio Chain

Unless the current AM studio is outfitted with stereo equipment, it will be necessary to install new stereo sources such as CD players, consoles and other distribution facilities. If good engineering practices are employed during the installation and layout phases, this portion of the process should not present major problems.

One area to be concerned about, however, is to ensure proper audio phase and amplitude matching of left and right channels throughout the facility. Without adequate tracking between both channels, proper monophonic frequency response and stereo imaging will not be maintained.

Figure 4.2-1 shows the resulting loss in monophonic frequency response due to improper phase relationship between the two audio channels. Another consequence of poor channel matching is a rapid degradation in L + R (monophonic) to L − R (stereo), and L − R to L + R crosstalk. This parameter typically degrades with frequency, reducing high frequency monophonic coverage and altering the high frequency stereo image.

Figure 4.2-2 graphically shows the degree of amplitude and phase matching required to obtain an arbitrary amount of crosstalk. For instance, a 0.1 dB amplitude error combined with 1° of phase mismatch will limit L + R to L − R and L − R to L + R crosstalk to about 40 dB. This relationship holds true for stereo separation during transmitter equalization.

AM Stereo Processing

Due to the nature of AM stereo transmission, the audio paths are different than its FM multiplex counterpart. Rather than operating in a discrete left and right channel mode, AM stereo transmission necessitates the use of matrix mode for transmission, since the amplitude modulation information comes from the combination of L + R (monophonic) audio and the RF carrier employs some form of phase modulation representative of the L − R (stereo) information. Unlike composite FM stereo, these paths are independent of each other and may also be processed independently. In fact, since the two audio paths are combined in matrix form before the actual modulation occurs, processing for AM stereo is most effective if it is also done in matrix form.

Figure 4.2-3 is a representation of a common and very effective tool used in AM stereo installation and maintenance: the X − Y or Lissajous pattern of a two channel oscilloscope. Standard operation is with the X input driven by the left channel, and the Y input driven by the right. This will produce a straight line display of 45° in the first and third quadrants for total L + R modulation, and a straight line display of 45° in the second and fourth quadrants for total L − R information.

If conventional left and right channel processing is used, the limiting must be done symmetrically, and at a level equivalent to 50% L + R and L − R modulation if only one channel were being applied. In this manner, proper 100% limiting will occur when L = R for the envelope (amplitude) modulation, and when L = (−R) for the phase modulation. However, under single channel conditions, there is a 6 dB loss in both monaural loudness and coverage. This undesired effect can be improved if the processing is done in matrix form.

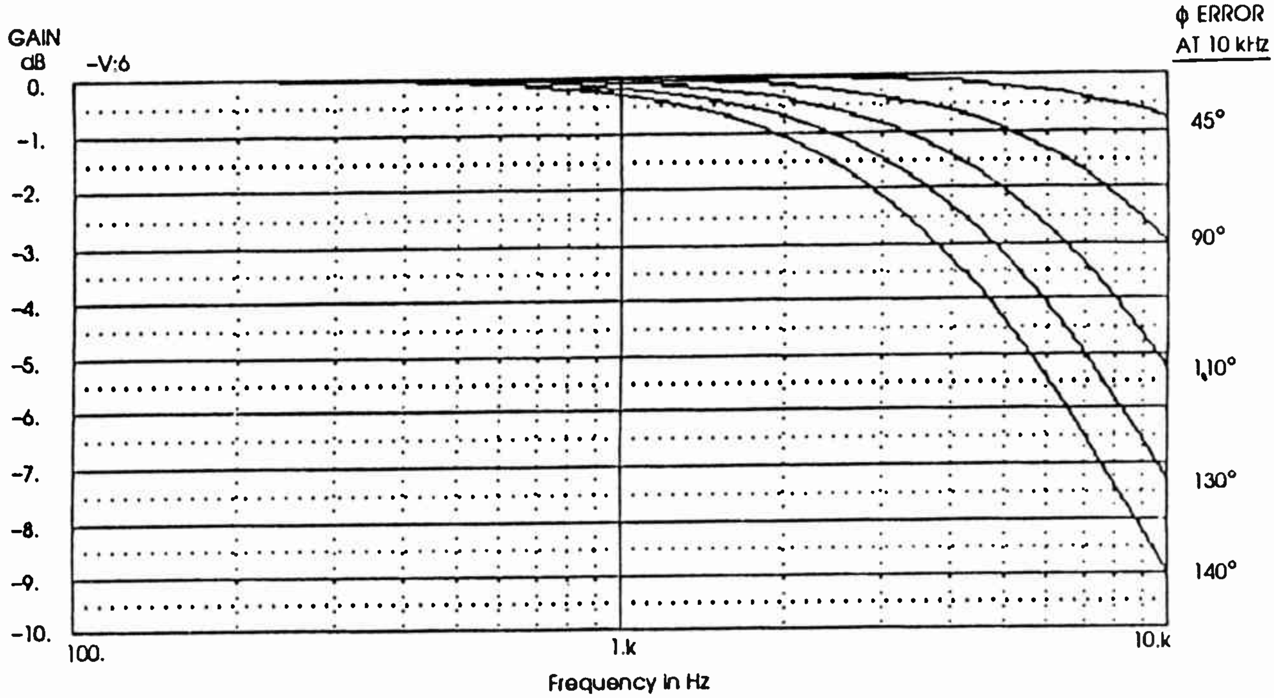
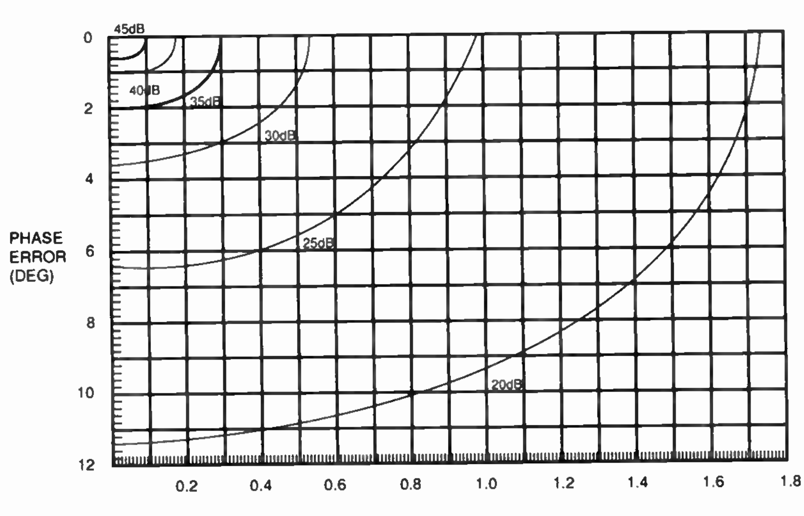


Figure 4.2-1. Loss in monaural frequency response due to phase errors.

Figure 4.2-4 shows the classic diamond shape produced by full matrix stereo processing. With this method, left and right channels are increased under single channel conditions, and under heavy stereo conditions, to maintain full envelope modulation. This mode of operation is called *full monaural support matrix stereo limiting.*

Unfortunately, instead of the 6 dB loss in monaural loudness associated with the conventional processing method, the full matrix method results in a 6 dB *increase* in stereo single channel loudness. However, this has been found to be less objectionable than the alternative, and combined with the infrequent nature of single channel conditions, is a more desirable side effect for AM stereo.

There is still one aspect of full monaural support matrix stereo limiting that makes it unacceptable. The unique decoder requirements necessary for the C-QUAM AM stereo system necessitate that the left and right channels be limited to -75%, where -100% is



General Equation:

$$\text{Separation }(A,\theta) = \left[\frac{(\cos\theta + A)^2 + (\sin\theta)^2}{(\cos\theta - A)^2 + (\sin\theta)^2}\right]$$

Where: $A = \frac{L-R}{L+R}$ Amplitude ratio

$\theta = \frac{L-R}{L+R}$ Phase error in degrees

Figure 4.2-2. Crosstalk (separation) versus amplitude and phase errors in a matrix system.

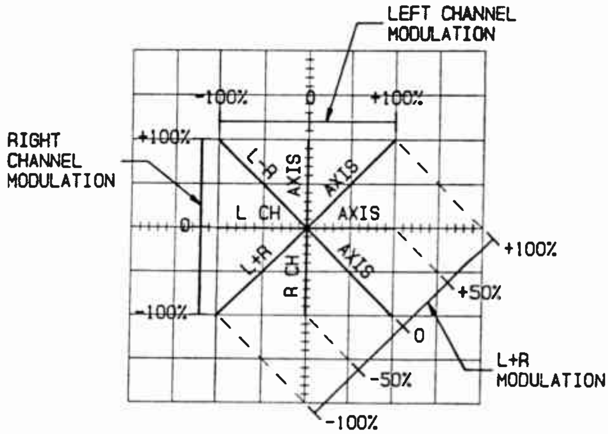


Figure 4.2-3. Conventional "X – Y" oscilloscope display.

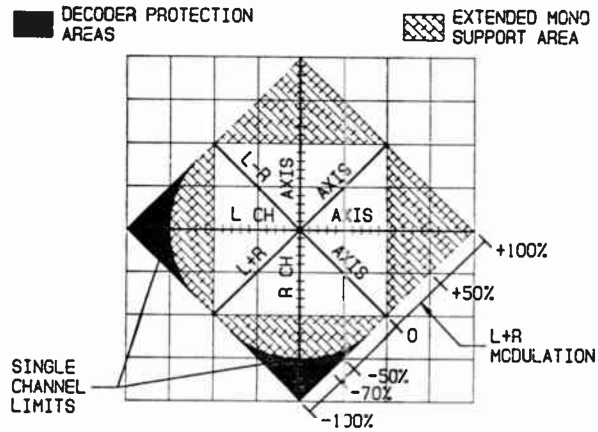


Figure 4.2-5. Modified matrix stereo processing.

equivalent to full envelope modulation caused by a single channel input. Therefore, all professional AM stereo matrix processors include a single channel limiter to prevent these problems. The new operational area is shown in Figure 4.2-5. Notice the protection areas that have been included to prevent non linear stereo operation. This style of AM stereo processing has been referred to as *modified* monaural support matrix stereo limiting.

Proper alignment and operation of any matrix AM stereo processor requires a thorough understanding of these departures from conventional FM stereo processing, as well as the practical limitations imposed by current AM stereo hardware designs. Without a proper grasp on these concepts, the processing can do more harm than good, and has often been the source of great frustration during installation and maintenance.[1]

Studio to Transmitter Link

Many AM stations operate studios at locations other than that of the transmitter. Therefore, there must be some form of studio to transmitter link (STL) to carry

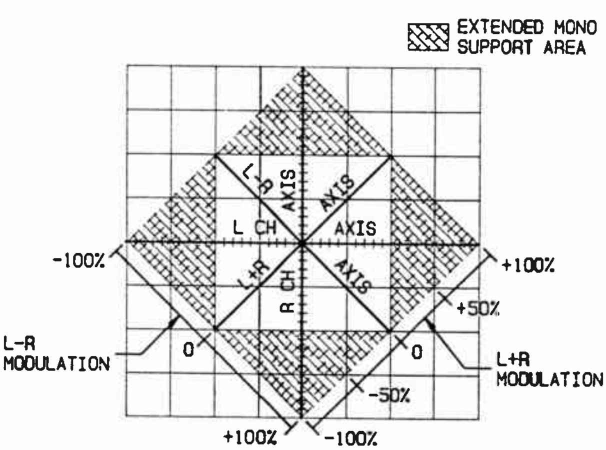


Figure 4.2-4. Full matrix stereo processing.

the audio information to the transmission facility. Traditionally, this has taken one of two forms: equalized phone lines or a microwave radio link.

Stereo transmission requires a second link to be installed. If this is not done carefully it can be a serious limitation to proper stereo operation.

Phone Lines

Installing a second phone line is often the first choice for AM stations that are currently using one. Two of the more common complaints with this approach are the increase in monthly fees associated with the second line and the troublesome problem of maintaining the amplitude and phase matching between the two lines.

Equalized phone lines typically have complex filters and frequency shaping equalizers. This makes them difficult to match and also prone to drift with time and temperature. The complexity of the circuits can also increase the total possible phase shift, making complete monophonic cancellation, or *combing*, a problem under extreme conditions. This same phenomenon will also cause complete mono to stereo image rotation, that will cause unexpected stereo imaging and excessive phase modulation.

Unfortunately, these paths are not under the direct control of the station, and one is often left to the mercy of the local telephone company for support. See Chapter 3.11, "Common Carrier Audio Program Services" for more information on current telco-based service options.

Radio Links

Installing a new radio link is generally considered a more favorable alternative, providing the frequency allocation is available. It also relieves the monthly financial burden associated with rented phone lines. However, discrete left and right STL transmission can suffer some of the same problems with amplitude and phase matching, but usually not to the same degree as phone lines. The STL transmitters and receivers will have audio, RF and IF filters, and depending on the design, construction and alignment, may not be well matched. It is best to consult the manufacturer to get an assurance that the two systems will track adequately.

There is a third approach that has been used in some locations with great success. An FM stereo generator can be used to create a baseband composite FM stereo signal that can then be transmitted on a single wideband STL link. The signal is then decoded back to left and right at the transmitter location using a high quality stereo decoder. This approach can be more expensive than discrete radio links, but results in a very high quality, reliable and stable way to provide the two channels to the transmitter.

Transmitter Preparation

Now that the studios are stereo ready, and the two channels are present at the transmitter site, the real job of installing AM stereo begins. Depending on the age and model of the current transmitter, preparing it to accept AM stereo can range from minor modifications to the nearly impossible.

Unfortunately, there are no absolutely defined procedures for making a transmitter stereo ready. It is highly recommended that a competent and reliable broadcast consultant be used to prepare the transmission chain. A reputable consultant with AM stereo installation experience can save time, trouble, and in the long run, money, since the consultant will know what is required for your unique set of circumstances and can implement them correctly.

In general, the work that will be required can be broken down into a few basic categories.

General Maintenance

The first step to high quality AM stereo is to have high quality AM monaural. Many AM transmitters have been neglected for years. A monaural proof-of-performance will show if the system is up to specifications. If not, it needs to be fixed so that it produces a decent proof. This is definitely money well spent, since it will improve your on air sound to all listeners, both monaural and stereo. It will also make the stereo installation easier, since there will be enough other problems to compensate for without the added burden of poor monaural performance.

Factory Modification Kits

Most manufacturers of AM transmitters will provide support to help upgrade their transmitters for AM stereo. Past experience has shown them where the stereo problems will be for any given model. In addition, some manufacturers have standard modification kits available, with instructions, to prepare the transmitter for stereo.

IPM Reduction

Incidental phase modulation (IPM) is broadly defined as any undesired angular phase shift of the RF carrier. It can be further broken down into two sub categories: IPM caused by power supply ripple (or induced magnetic fields), and IPM caused by envelope modulation.

The first form of IPM most often results from insufficient power supply bypassing. Before the advent of AM stereo, transmitter manufacturers were not concerned about IPM, since it was not decoded by conventional envelope detectors and generally did not affect the AM signal. It was simply not worth the additional cost to provide the extra bypassing.

Its main effect is to limit the decoded L − R signal-to-noise ratio (S/N), and is often the dominant stereo noise component. Any improvement in this area will improve the decoded stereo S/N. It should be possible to reduce this form of IPM to −45 to −55 dB below 100% L − R modulation.

The second form of IPM is much more troublesome to reduce and more detrimental to quality AM stereo. This mode of IPM is caused by amplitude modulation of the carrier, and can be caused by several mechanisms. For tube type transmitters, it is often the result of poor neutralization of the final PA stage, and can result in equivalent L − R modulation levels in excess of 25%.

The interelectrode capacitances found in vacuum tubes can cause instabilities. In particular, the plate to grid feedback capacitance provides a path for positive feedback. Figure 4.2-6 shows the static interelectrode capacitances found in a tetrode. The C_{g-c} and C_{g-p} are the two capacitances of importance for proper neutralization. The intent of neutralization is to cancel the C_{g-p}, so as to minimize its effect, both on stability and on IPM. The most popular method of neutralization is the *capacitance bridge*. Figure 4.2-7 shows a simplified tetrode configuration employing classic bridge neutralization. C_{g-p} and C_{g-c} are the grid to plate and grid to cathode interelectrode capacitances, respectively. C is comprised of any input capacitance, including stray capacitance, and is required for complete neutralization of C_{g-c}. C_n is the neutralization capaci-

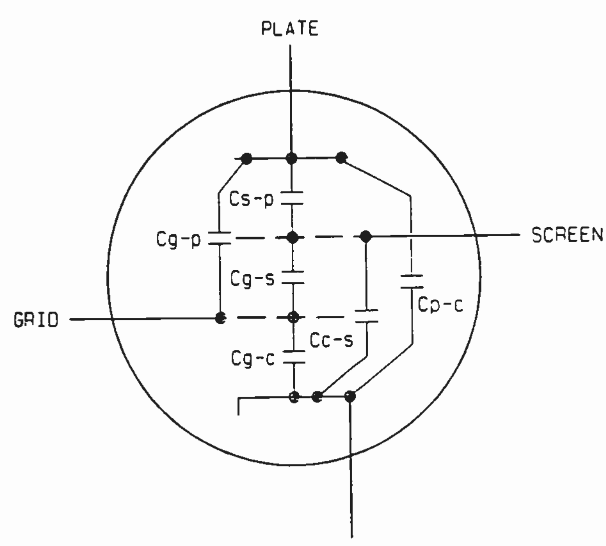


Figure 4.2-6. Static interelectrode capacitances of a tetrode tube.

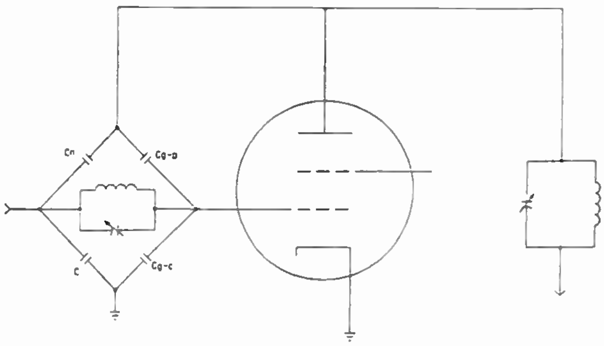


Figure 4.2-7. Bridge neutralization.

tance required to balance C_{g-p}. For proper balance (proper neutralization), the capacitor ratios must satisfy the following relationship:

$$\frac{C_n}{C} = \frac{C_{g-p}}{C_{g-c}}$$

With proper neutralization, IPM levels should be able to be reduced, typically, to -35 to -45 dB below 100% L − R modulation.[2,3]

For the solid-state transmitter, IPM is most often caused by the nonlinear output capacitance of the solid state device. The value of this capacitance is a function of the applied collector or drain voltage. Since most solid state transmitters use high level modulation, this voltage varies at the audio rate, that will affect the RF phase angle.

Fortunately, most solid state transmitters are relatively recent designs and steps have been taken to reduce this effect to an acceptable level. Many new transmitters are shipped stereo ready, requiring no additional effort to reduce IPM.

Tuned Circuits

Some older transmitters used narrow bandwidth tuned RF circuits as input, interstage and output coupling networks. Proper stereo operation requires that the amplitude and phase relationship of the L + R (envelope) information match that of the L − R (phase modulated) information. These narrowband tuned networks have three undesired effects. First, they introduce a non-uniform time delay to the RF carrier, resulting in complex phase equalization problems. Second, they result in L − R high frequency response problems, further aggravating the equalization requirements. Finally, the narrowband nature of the network alters the sideband structure of the phase modulation, resulting in increased distortion.

It is a fairly common practice to increase the bandwidths of these circuits to improve the stereo signal. A word of warning, however: reducing the Q of a tuned stage will result in a decrease in voltage at the output of the network. Care must be taken to ensure the resulting signal is adequate to drive the following stage. It is wise to consult the factory or a knowledgeable consultant before attempting this modification.

Antenna and Phasor Alignment

All AM stereo transmissions are sensitive to amplitude and phase non-linearities to one degree or another. For best performance, the antenna system should be as broadband as possible, and as symmetrical as possible. The C-QUAM system especially requires good transmission bandwidth and phase linearity.

If required, a consultant should be retained to broadband the antenna system and ensure its fitness for AM stereo transmission.

Monaural Proof of Performance

One final step should be carried out before the actual stereo conversion begins. A complete monaural proof of performance should be done to ensure the system is ready to accept stereo, and as a record to compare final monaural performance after installation.

STEREO EXCITER INSTALLATION

With a properly operating monaural transmission chain in place, the actual transmitter conversion to AM stereo may begin. Mount the AM stereo exciter close enough to the transmitter to allow proper interconnection. A general rule is the RF interconnection cable should be no more than 30 ft from the transmitter.

RF Interfacing

Since AM stereo transmission uses some form of phase modulation of the RF carrier, the AM stereo exciter will have a phase modulated RF output that will replace the transmitter's internal RF oscillator. The rated output power varies from exciter to exciter, from a few watts up to 10 W.

It will be necessary to pick a suitable RF insertion point to break the internal RF chain and replace it with the RF from the exciter. Stereo-ready transmitters will be outfitted with a BNC connector, usually requiring a transistor-transistor logic (TTL) level (5 V, peak-to-peak) signal for proper operation. Other transmitters require modification to provide a way into the RF sections.

If the particular type of transmitter uses narrow tuned RF stages, it is a good idea to pick the farthest point in the transmitter that the exciter can still provide adequate RF drive.

Many transmitters can accept a TTL level signal but are not outfitted with the necessary interface. Figure 4.2-8 shows a general purpose interfacing network that may be used in these transmitters.

Providing the Proper Termination

The RF output of an AM stereo exciter is designed to operate in a 50 Ω system. This is required due to the extended distances from the RF source to the transmitter. For proper operation, the transmitter end of the 50 Ω coax ***must*** be terminated with 50 Ω to prevent reflections on the line. This is especially important if the RF output is a square wave, which it is most likely to be. Improper termination will cause reflections on the line, that can cause excessive ringing

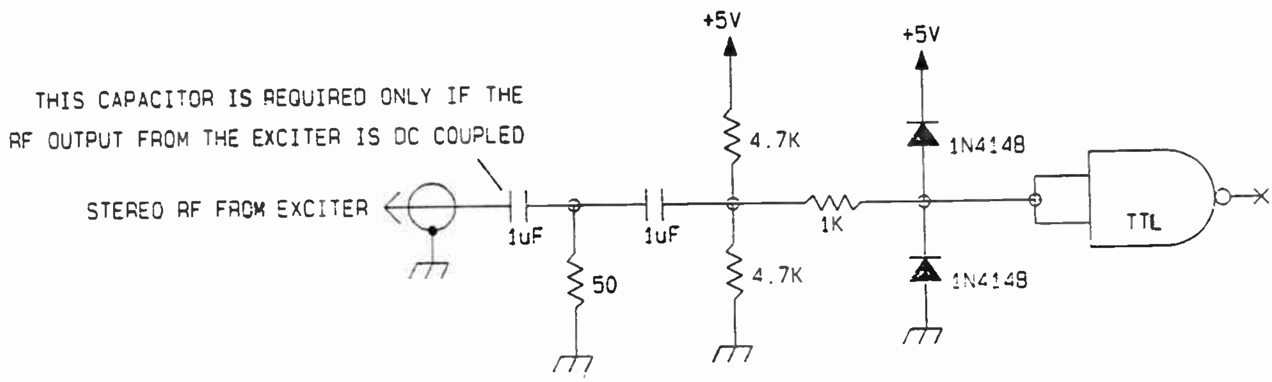


Figure 4.2-8. TTL RF interfacing.

and, for TTL inputs, improper logic operation. If the ringing is excessive, it is possible to damage the input stage due to excessive voltages. For TTL compatible input levels, a 1/4 W resistor is adequate, provided that the input to the transmitter is ac coupled, then dc restored (see Figure 4.2-8). If higher power levels are required, select the proper resistor power rating based on the following formulas:

$$P_{diss} = \frac{(V_{rms})^2}{50} \quad \text{(for sine wave input)}$$

$$P_{diss} = \frac{(V_{peak})^2}{50} \quad \text{(for square wave input)}$$

Internal RF Oscillator

Since the transmitter's original RF oscillator has been replaced by the AM stereo exciter, some possible problems may need to be dealt with. First, there may be some coupling between the original oscillator and the new RF signal. This coupling manifests itself most often in the form of a beat frequency equal to the difference in the two oscillators frequency calibration.

The most common cure is to simply remove the original crystal from the transmitter, or disable the power supply to the oscillator. This will prevent beat frequencies, and is a minor enough modification to allow quick replacement should the original oscillator be needed in an emergency.

If the old oscillator should be required due to a failure in the AM stereo exciter, it would be nice to be able to quickly switch back to the original oscillator and return to the air, in monaural, until the problem can be resolved. This can be done fairly easily with a relay. Figure 4.2-9 shows one possible implementation of an RF switching circuit. Notice the extra relay contact used to remove the power from the original oscilla-

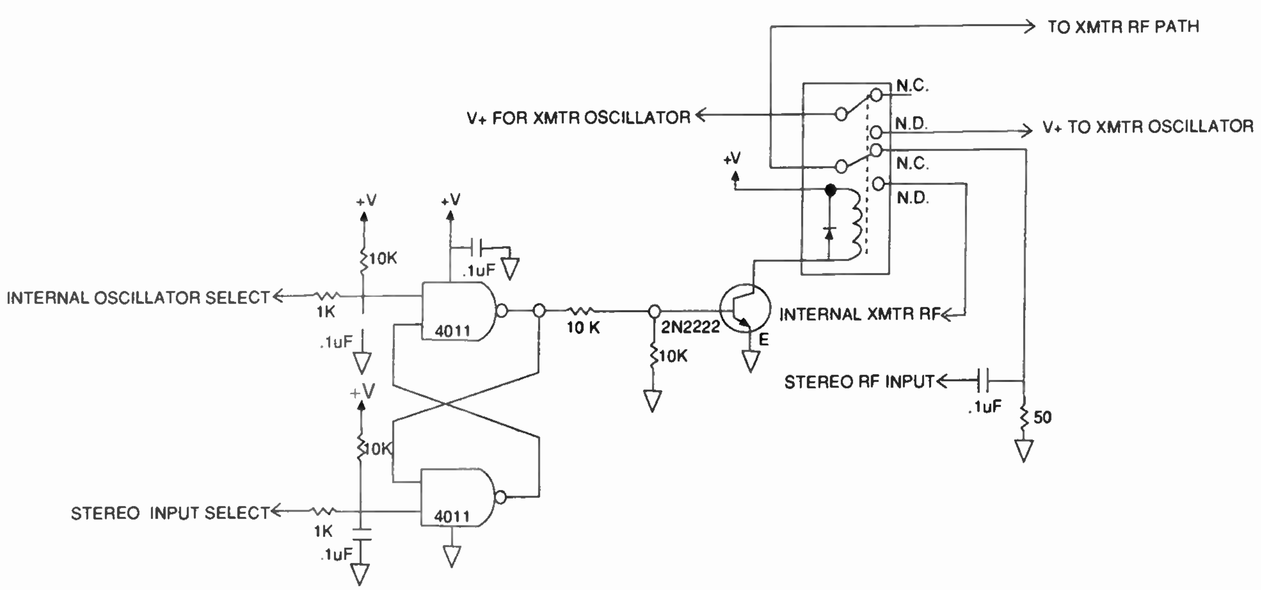


Figure 4.2-9. Mono/stereo emergency switching.

tor during normal stereo operation. Again, this eliminates the possibility of beat frequencies.

Audio Interfacing

The left and right channels must be connected to the exciter, and the levels set according to the manufacturers recommendations. The existing monaural input to the transmitter should be replaced by the monaural output from the AM stereo exciter. All audio connections should be made using a high quality shielded cable, such as Belden 8451. This is important to prevent RF pickup on the cable. To prevent ground loops, ground the shield at only one end.

AM STEREO EQUALIZATION

Now that all the necessary pieces are in place, the final part of the AM stereo installation can begin. Since half of the stereo information is contained in the monaural (L + R) connection to the transmitter, and the other half is in the form of phase modulated RF (L − R), they must arrive at the modulator portion of the transmitter at the same time. This is required so that proper left and right channels may be recovered in the receiver after de-matrixing. The amount of amplitude and phase (delay) matching required at the transmitter's modulator to produce a given amount of stereo separation after de-matrixing can be seen in Figure 4.2-2.

Without proper audio equalization, both stereo separation and distortion would suffer. Unfortunately, the amount of delay and amplitude equalization required for one type of transmitter is different from any other. In addition, the circuits in the transmitter that cause the problems are generally complex, so the best that any AM stereo exciter equalizer can do is to approximate the response shape. The effectiveness of the equalizer is dependant on the type of transmitter and the complexity of the antenna pattern at any given station.

For these reasons it is clear why there are no firm rules to follow during equalization. It is often a trial and error routine, especially at the higher audio frequencies where the response shape is most complex. If an AM stereo exciter was designed specifically for one brand of transmitter, it would be possible to equalize the two paths almost perfectly, but since it is a generic conversion tool, it can, at best, only come close.

Types of Equalization

There are three different equalization sections in an AM stereo exciter. Each is intended to fix a particular frequency range. Some installations require all three, others only two, and very rarely, only one will be needed.

Group Delay

The first section which is almost universally required is the *group delay equalizer*. Its purpose is to provide a constant time delay to either the L + R or L − R paths, depending on which one arrives at the modulator first. It is primarily used to equalize the low to mid frequencies (up to a few kilohertz). If the response shape of the offending path in the transmitter is low-pass (or bandpass for the RF path), the time delay for the lower frequencies will be constant, so the alternate path can include this equalizer to compensate adequately.

A common example would be any AM transmitter employing pulse width modulation (PWM), in that case the L + R path would include a PWM low pass filter in the modulator just ahead of the RF amplifier. The L + R audio will have significantly more delay than the L − R (RF phase modulation) path. In this case, the group delay equalizer would be added to the L − R path to slow it down so that it arrives at the RF amplifier at the same time the L + R signal does.

On the other hand, some transmitters will have more low/mid frequency delay in the L − R path due to several RF interstage bandpass filters. In this instance, the group delay equalizer would be added to the L + R path to compensate.

Low Frequency Equalization

The second type of equalizer found in the exciter is the least often used section. It is only necessary if the low frequency response of the L + R path is non-uniform. This is found primarily in older plate modulated transmitters where the reactance of the plate transformer causes phase, and to a lesser degree, amplitude nonlinearities.

Some transmitters will also include an active audio high-pass filter in the monaural input. This can also be effectively corrected by the low frequency equalizer.

High Frequency Equalization

The third audio equalizer is designed to approximately complement the complex high frequency response shape of most AM transmitters. The amount and shape of high frequency correction varies greatly from transmitter to transmitter, that makes this section the most difficult adjustment of the installation process.

Since this section is only meant to approximate the high frequency amplitude and delay characteristics, it will not correct at all frequencies. It is not uncommon to be caught in a seemingly endless loop of adjusting for one frequency, only to find out you have made another frequency worse. The key to a successful installation is the ability to recognize a good overall compromise and once you have quit. This is one area where a good engineering consultant with AM stereo installation practice will be helpful.

Figure 4.2-10 shows the various equalization adjustments found on one C-QUAM AM stereo exciter. Notice the additional set of equalization labeled *Night*. This is useful for stations utilizing a different transmitter, antenna pattern or both for part of the broadcast day. It can also be used to equalize an emergency backup transmitter.

Equalization Path Selection

The proper alignment of the three equalization sections is the key to high quality AM stereo. However, before they can be adjusted, they must be inserted into

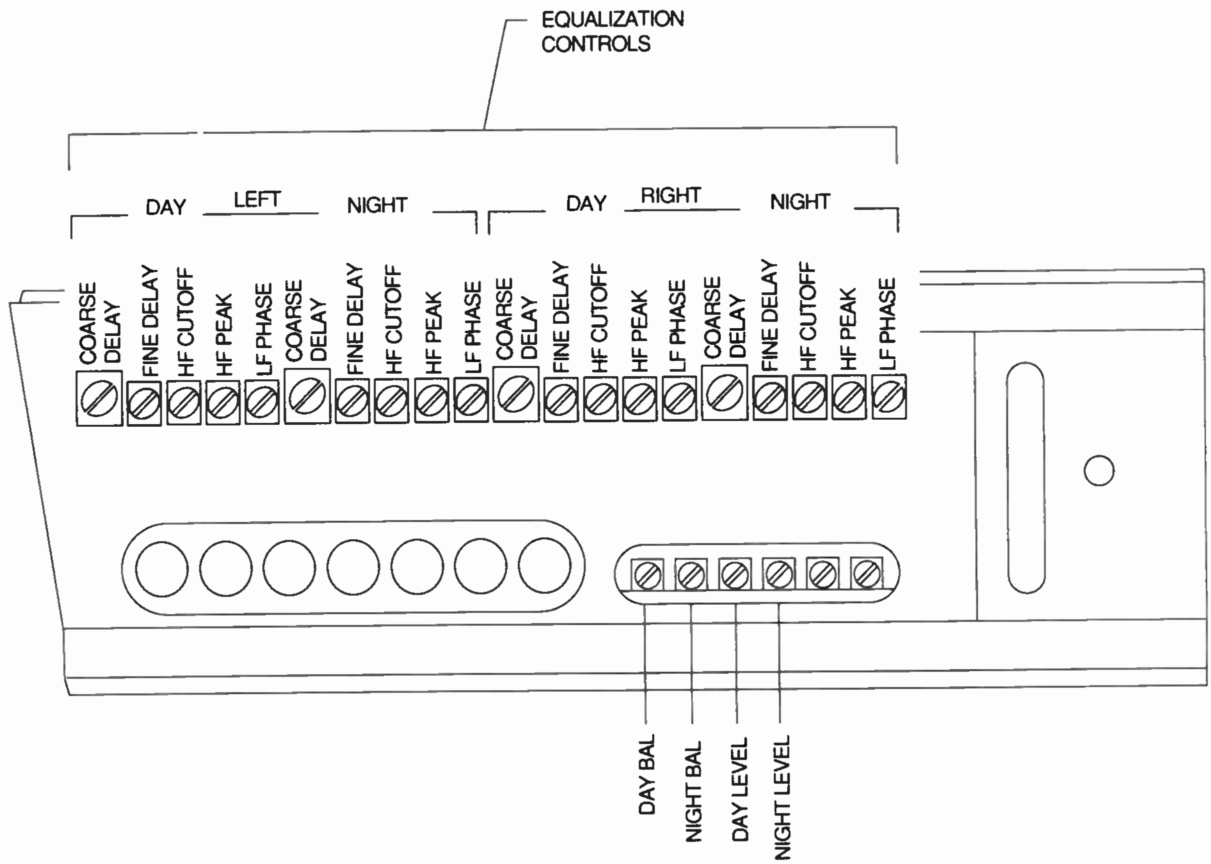


Figure 4.2-10. For a stereo exciter-stereo equalization control points.

either the L + R path or the L − R path, depending on where they are needed.

For someone with an intimate knowledge of the design of a particular transmitter, it may be possible to know for sure which path needs which equalizer. However, it is also possible that unique antenna characteristics could change what would be necessary into a 50 Ω resistive load. What results is a *try it and see* method. It may be necessary to change paths during the installation procedure as other equalizer sections are included. This will happen when, for instance, the inclusion of the high frequency equalizer in series with the group delay equalizer actually forces the constant delay to be added to the other path to compensate for low and mid frequency delay found in the high frequency equalization.

Figure 4.2-11 shows one method of selecting the proper path for any equalizer. This matrix switch is a flexible patch bay approach to inserting equalizers in series or parallel into either the L + R or L − R paths.

Equalization Procedure

Figure 4.2-12 outlines a typical test setup required for AM stereo equalization. Test equipment required for an adequate installation can be as simple as a low distortion audio oscillator/analyzer, an AM stereo modulation monitor and a dual trace oscilloscope with X − Y capability. In addition, a spectrum analyzer capable of resolving audio sidebands at the AM band frequencies can be very useful, but is not mandatory.

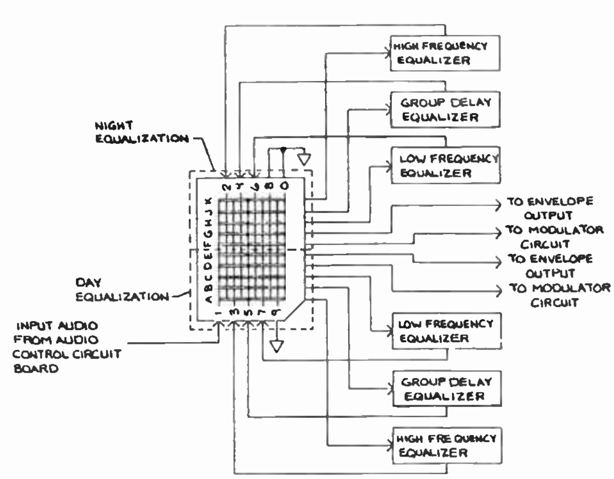


Figure 4.2-11. Stereo generator matrix switch.

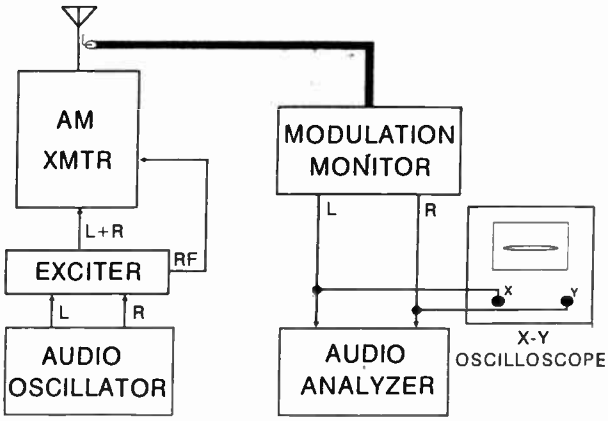


Figure 4.2-12. Stereo equalization test setup.

The X − Y Oscilloscope

The fundamental problem in AM stereo is determining the type and path of equalization required. In addition, it is necessary to determine if the system requires phase or amplitude equalization. A dual trace oscilloscope operating in the X − Y or Lissajous pattern mode is capable of resolving between amplitude and phase errors (or a combination thereof), as well as showing a multitude of additional information such as harmonic distortion, IPM, negative limiting, modulation compression and so on. To the trained eye, it is one of the most useful tools in AM stereo installation. Without this tool, it would be almost impossible to equalize the system.

AM Stereo Modulation Monitor

The final performance results of a stereo installation will only be as good as the equipment used to measure them. One of the most critical pieces of test equipment for proper installation, and just as important, routine maintenance, is the *AM stereo modulation monitor.* It should be a high quality piece of test equipment, and simple to use.

An Equalization Example

AM stereo exciters will have similar equalization adjustments, but may be labeled differently. The *day* adjustments are used in this example, though the *night* procedure is identical.

The first adjustment to be made is to set the exciter's L + R, or envelope, level to get the correct amplitude modulation for a corresponding L = R input. This is, at this stage, only a coarse adjustment. It can be done at any convenient modulation percentage. For instance, input 1 kHz L = R to the exciter such that the L + R meter on the exciter reads 80%. Then adjust the envelope output level of the exciter until the transmitter is modulating 80% as measured on the modulation monitor.

Next, apply a 1 kHz tone to the input of the left channel, only, at a level equivalent to 50% amplitude modulation. The oscilloscope should show a horizontal line (assuming the "X" input is the left channel). Most likely, it will be neither exactly on the X axis, nor a perfectly straight line. Figure 4.2-13(a) shows a typical display. If a display similar to Figure 4.2-13(b) is seen, the polarity of the L + R connection to the transmitter is wrong. This can easily be fixed by reversing the plus and minus terminals at the transmitter input. Figure 4.2-13(b) actually shows the high second harmonic distortion produced when the polarities of the amplitude modulation and phase modulation are not correct.

Next, the fine adjustments to maximize separation can be made. While monitoring both the residual right channel separation and the oscilloscope, insert the left channel group delay in either the L + R or L − R path (Figure 4.2-11) and adjust the coarse group delay. If the display opens up rather than closes, and the separation worsens, it has been inserted in the wrong path (see Figure 4.2-14). Reverse the location and adjust the coarse and fine group delay until the display closes (Figure 4.2-15). Then, fine tune the envelope level until the display lies exactly on the X axis and the left to right channel separation nulls (Figure 4.2-16).

Apply a 1 kHz tone to the input of the right channel,

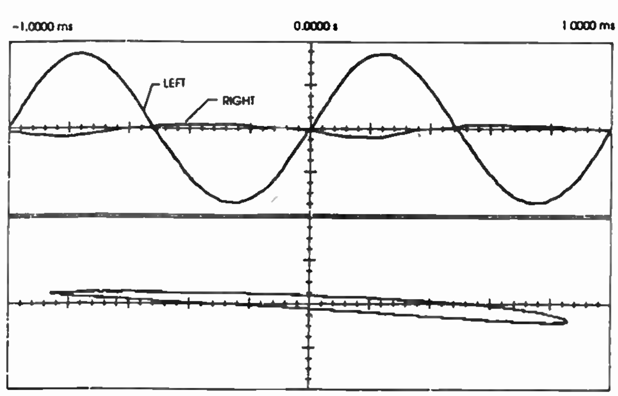


Figure 4.2-13(a). Typical left channel only display before equalization.

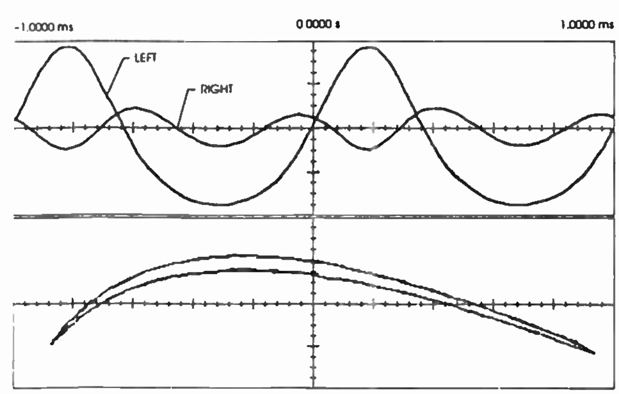


Figure 4.2-13(b). X − Y display caused by improper L + R polarity at transmitter.

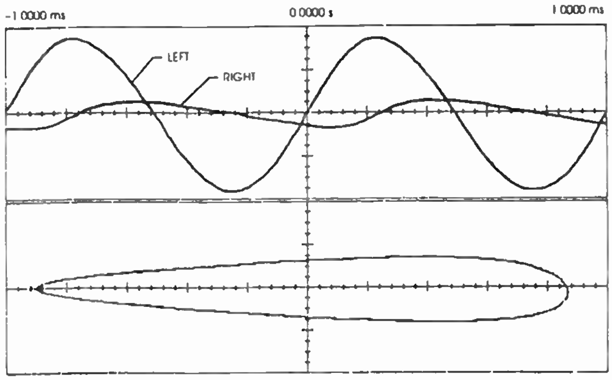


Figure 4.2-14. Left channel only with phase equalization error.

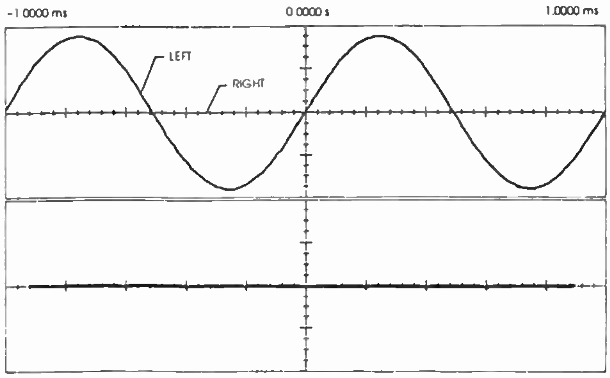


Figure 4.2-16. Left channel only, properly equalized.

only, at a level equivalent to 50% amplitude modulation. The oscilloscope should show a vertical display with only phase error. Insert the right channel group delay in the same path as the left channel and adjust until the display closes. The residual separation in the left channel will also null. Theoretically, the envelope level adjustment should not have to be adjusted for the right channel, but often the right to left separation can be improved slightly by fine tuning the level, at the expense of left to right separation. It is best to compromise the two to obtain the same amount of separation left to right as right to left. Some exciters are equipped with a balance control to help this problem. Refer to the individual service manual for the correct adjustment procedure.

The next step is to check the low frequency separation. A good frequency to look at is 100 Hz. If the separation has degraded significantly, the low frequency equalizer can be used to correct the problem. Again, this is usually only necessary in plate modulated transmitters, or ones that employ low level audio input high pass filters. The procedure is very similar to the group delay adjustment.

If low frequency equalization is used, re-check separation at 1 kHz and adjust the group delay, if necessary, to compensate.

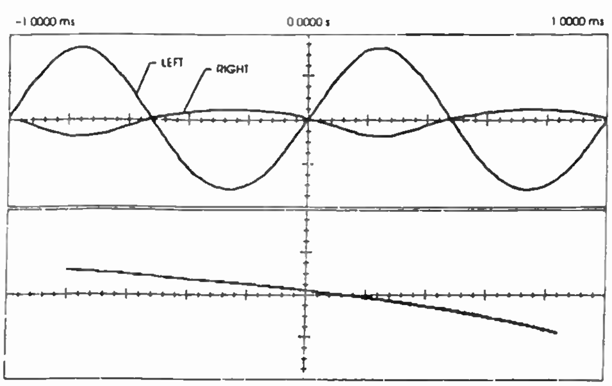


Figure 4.2-15. Left channel only with amplitude equalization error.

At this point, the transmitter should be equalized for good separation (30–40 dB) from 50 Hz to a few kilohertz. A quick check of separation at 5–10 kHz will most likely show a rapid degradation in separation performance.

With the left channel high frequency equalization controls set at minimum, insert the H.F. equalizer in one of the paths. Input 7.5 kHz left channel at a level sufficient to produce 50% envelope modulation. Observe the oscilloscope display while adjusting the H.F. cutoff control. If the phase degrades (display opens), then switch the paths. If it improves, adjust until it closes. Input 1 kHz and re-adjust the group delay to compensate for the added mid-frequency delay, and repeat until 1 kHz and 7.5 kHz are equalized for phase.

If the display at 7.5 kHz is not on the X axis, adjust the H.F. peaking and then the H.F. cutoff until the display lies on the X axis with minimum phase error. Again, adjust the group delay control at 1 kHz to compensate and repeat the procedure until maximum separation is achieved at both 1 kHz and 7.5 kHz. Repeat the procedure for the right channel.

Once this is done, spot check the separation performance from 1–10 kHz. Using good judgement, it may be necessary to compromise the performance at 7.5 kHz to improve it elsewhere. The procedure is identical with the exception of X kHz in place of 7.5 kHz. While there is no such thing as a typical proof, it should be possible to achieve greater than 20 dB separation out to 10 kHz.

The Stereo Proof

When satisfactory separation performance is obtained, the final step in the AM stereo installation is to run a full proof-of-performance, both monaural and stereo. In particular, pay attention to high frequency distortion. It may be necessary to fine tune the equalization to reduce distortion at the expense of a few dB of separation. An acceptable proof should show less than 3% THD and greater than 20 dB separation at all frequencies. This proof is a valuable tool during routine maintenance to gauge performance and to assess whether the equalization needs adjusting.

ROUTINE MAINTENANCE

Any change in the transmission system amplitude or phase response will cause a degradation in stereo performance. These changes will occur for several reasons including: seasonal environment changes, routine transmitter tuning and vacuum tube variations over its lifetime.

It is a good practice to include a spot check of stereo performance as a routine maintenance item. This should be done at an interval of no more than every six months. Left unattended, these normal system variations can cause a serious degradation in audio quality, but a small time investment will insure continued high quality AM stereo performance.

SUMMARY

Properly installed, AM stereo can offer many advantages to the AM broadcaster with the ever increasing pressure for high fidelity audio. Good planning, execution and follow up to the installation, are essential to avoiding the creation of new problems.

REFERENCES

1. *Stereo Processing,* Circuit Research Labs, Inc., Tempe, AZ, 1983.
2. Stanley, James S., *Engineering Report: Wilkinson AM-1000B Stereo Conversion,* Stanley Broadcast Engineering, Tempe, AZ, 1987.
3. Orr, William I., *Radio Handbook,* 20th Edition, H.W. Sams & Co., 1975.
4. Orban, Robert, and Oganowski, Greg, "AM Broadcast and Receiver Standards—Towards Opening a Dialogue," *SAE Technical Paper Series,* Warrendale, PA, 1985.
5. *Electronic Circuit Analysis 2.50 MS-DOS Version,* Copyright 1990 by Tatum Labs, Inc, Ann Arbor, MI.
6. Anthony, Edward J., *Second Generation Techniques for AM Stereo Exciter Design,* Broadcast Electronics, Inc., Quincy, IL, 1984.

APPENDIX A
MOTOROLA C-QUAM® AM STEREO SYSTEM

Greg Buchwald
Motorola Inc., Schaumburg, IL

C-QUAM SYSTEM EQUATIONS

Any broadcast signal can be broken into three major components: amplitude, frequency and phase. The equation for a monophonic transmission can be described by:

$$E_R = (1 + L + R) \cos (\omega_c t + \varphi)$$

where 1 represents the carrier, L + R represents the monophonic (or left and right) information to be sent, $\omega_c t$ represents the carrier frequency, and φ represents phase modulation information that is, ideally, zero.

In fact as one uses ±100% modulation of L + R as an example and substitutes ±1 into the equation, it is obvious that at L + R = +1, the carrier level is instantaneously twice as high and at L + R = −1, the carrier level is instantaneously 0.0, or cutoff has occurred. From this equation it is also obvious why negative modulation is limited to 100% while positive modulation can exceed 100%. In the United States, the positive limit is 125% or 1.25 in the equation.

To insert stereophonic information, one could alter the amplitude, frequency or phase of the transmitted signal. Altering the amplitude is to be avoided since the envelope would no longer represent 1 + L + R, but instead, a distorted component containing L + R. Substantial alteration of the frequency is also to be avoided. This leaves only the phase component available for modulation.

One method of adding a second channel of information to an existing carrier and utilizing the same spectral assignment is to make use of Quadrature Modulation (QUAM) techniques. Linear QUAM, not unlike that used to convey the chroma information in NTSC color transmission, can be used to convey the second channel of information. Advantages of QUAM are:

- No increase in occupied bandwidth
- Little S/N degradation
- No loss of existing coverage
- The potential for synchronous reception techniques

However, the major disadvantage is that the envelope term is not 1 + L + R, but, instead is:

$$\sqrt{(1 + L + R)^2 + (L - R)^2}$$

leading to high levels of distortion in current monaural, envelope detector receivers and some difficulty in the conversion of existing broadcast transmitters due to additional requirements placed on the modulator stage.

The C-QUAM system was derived from QUAM, therefore it retains, to a large extent, the advantages of QUAM. In fact, C-QUAM can also use synchronous detection techniques particularly when conditions warrant. One difference between the generation of the C-QUAM signal versus the QUAM signal is found in the envelope audio term. By substituting the distorted term required by QUAM with the simple summation of the left and right channels, the envelope is made compatible with existing receivers. Mathematically, the system has been designed as follows.

The equation for the QUAM signal is:

$$E_R = \sqrt{(1 + L + R)^2 + (L - R)^2} \cos (\omega_c t + \phi)$$

$$\text{where } \phi = \tan^{-1}\left[\frac{L - R}{1 + L + R}\right]$$

The desired envelope component is: $E = (1 + L + R)$.

The cosine of the instantaneous phase modulation is:

$$\cos \phi = \frac{1 + L + R}{\sqrt{(1 + L + R)^2 + (L - R)^2}}$$

If QUAM is multiplied by cos φ:

$$\frac{1 + L + R}{\sqrt{(1 + L + R)^2 + (L - R)^2}} \times \sqrt{(1 + L + R)^2 + (L - R)^2} \cos(\omega_c t + \phi)$$

Then the resultant is:

$$E_R = (1 + L + R) \cos(\omega_c t + \phi)$$

$$\text{where } \phi = \tan^{-1}\left[\frac{L - R}{1 + L + R}\right]$$

In the process, the envelope term is made compatible by sending $1 + L + R$ while the in-phase (I) and quadrature (Q) components are multiplied by $\cos \phi$. Therefore the broadcast C-QUAM signal has the following characteristics:

Envelope (E) $= 1 + L + R$
In-phase (I) $= (1 + L + R) \cos \phi$
Quadrature (Q) $= (L - R) \cos \phi$

It can be seen that the mono, $1 + L + R$, signal may be directly derived from the envelope detector output, while a quadratine detector combined with division by $\cos \phi$ may be used to demodulate the $L - R$.

THE C-QUAM ENCODING SYSTEM

There are several forms by which the C-QUAM encoder may be implemented. In the first form, a series of linear, balanced multipliers is employed to generate a quadrature modulated signal. The signal is then amplified to the point of limiting in order to remove the QUAM envelope term. The summation of the left and right channels provides a distortion-free monophonic audio term which modulates the transmitter. The resultant phase modulated signal is used to replace the crystal oscillator stage of the transmitter. The existing broadcast transmitter conveys the L − R information in the form of complex phase modulation, and the envelope conveys the L + R information to both the existing monophonic as well as newer stereophonic receivers.

Other modulation techniques include the matrix switching method and the Time Division Multiplex method. The latter eliminates the need for audio matrices in the stereo modulator path.

To help visualize the generation of the signal, refer to Figure 4.2-A1, the block diagram for the exciter. Since the linear, balanced modulator technique is the easiest to understand, it will be described herein. Interested readers are encouraged to study literature produced by other C-QUAM broadcast equipment manufacturers to obtain a full understanding of the various modulation techniques.

Beginning with the RF stages, a crystal at four times the carrier frequency ($4F_c$) is fed to a digital divider circuit that results in an output that is an on-frequency carrier signal at 0°, 90°, 180° and 270°. These RF signals are utilized in pairs as differential references for three balanced modulators. The purpose of each balanced modulator is to suppress the carrier feedthrough and produce sideband information only when audio is present. It can be seen that the first two modulators are fed with the 0° and 180° reference signals. These modulators, therefore, produce an output referenced in-phase or synchronous to the carrier frequency. The third modulator is, conversely, fed with the 90° and 270° outputs from the RF divider and therefore forms a quadrature modulator stage. The first modulatory is fed with a dc voltage that causes a precise offset from the null of the 0° carrier reference, thereby producing a precise dc carrier term at its output. The second modulator is fed with L + R and therefore produces in-phase, L + R modulation sidebands. The final modulator is fed with L − R audio information and produces quadrature sidebands relating to the stereophonic information. It is important to note that the carrier-producing modulator is used in preference to simply unbalancing the in-phase (L + R) modulator for lower distortion.

The three balanced modulator outputs are summed and band-pass filtered to remove odd-ordered harmonics of the desired carrier frequency. Removal of odd order harmonics prior to limiting is important since odd ordered products represent a dc term that can unbalance the subsequent limiting stage, thereby introducing distortion. After band-pass filtering, the RF signal is amplitude limited by multiple stages that exhibit approximately 50 dB of limiting gain. The envelope component is therefore removed, leaving only the phase-modulated carrier component. This constant amplitude RF signal is then further amplified to a level adjustable from zero to 30 V, peak-to-peak. It is this amplified, constant amplitude carrier signal that is interfaced to the transmitter to create the composite stereo transmission.

Unlike FM stereo where a composite signal is fed into a wideband modulator input, the AM stereo signal is constructed at the power amplifier stage of the converted broadcast transmitter. A portion of the signal from the exciter enters the transmitter through an RF interface and follows one path to the RF power amplifier, while the remaining element, L + R, enters through the audio input terminals and follows a different path to the modulator where it is finally combined with the RF phase modulated information. Since two different circuit paths are used, the time delay along each path may be different, often by 40 μsec or greater. This level of delay can be understood when one considers the phase shift through the low pass filter in a conventional PDM transmitter that may exhibit a fifth or seventh order elliptical function. If the delay between the paths is not matched, both reduced separation and increased distortion will result. Loss of separation is easily understood since proper dematrixing of the

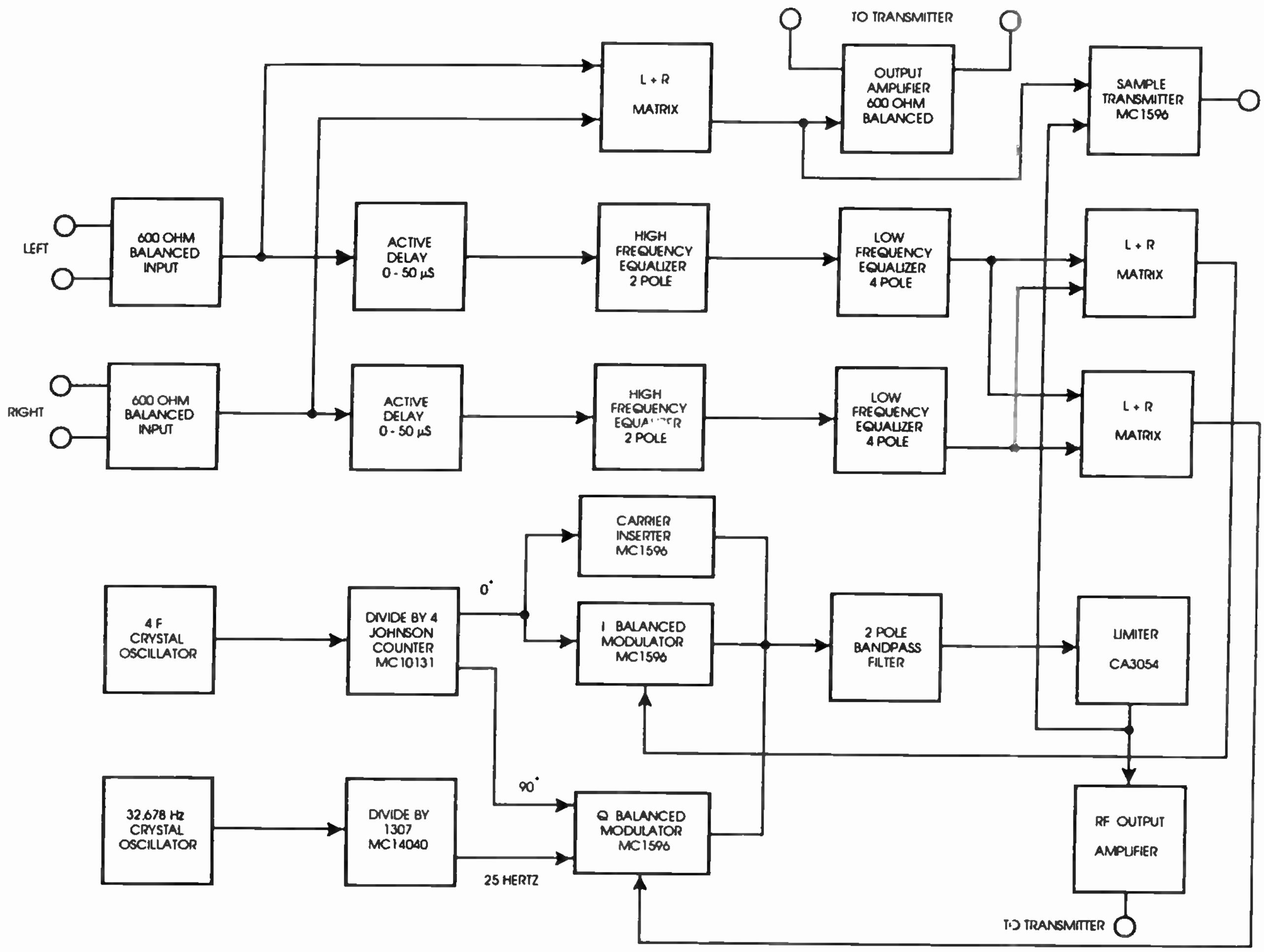


Figure 4.2-A1. Block diagram of C-QUAM exciter.

audio signals in the receiver will occur only if the phase and amplitude of L + R and L − R are closely matched. The increase in distortion is a concept that can only be clearly understood when one examines the phase modulated component of the signal and finds that an L + R term is found in the PM component. If this L + R term does not match the envelope term, an increase in distortion would result. Therefore, the audio equalizer is a very important section of the exciter.

In the simplest form of equalization, a delay circuit in the audio path to the quadrature modulator section of the exciter would be utilized, thereby introducing time delay into the L + R and L − R audio fed into the QUAM generator so that it matches the L + R audio delay through the broadcast transmitter. Although adequate results will occur with this approach, a modification of this approach allows for additional correction of the signal to compensate for bandwidth or sideband symmetry problems commonly associated with broadcast antenna systems, particularly directional arrays.

A once common form of single sideband generation particularly popular in the 1950s, utilized a scheme known as the *phasing method*. The process was simple because one could simultaneously amplitude and phase modulate a signal, essentially generating a QUAM signal. If the process is taken one step further so that both the RF terms and the audio terms fed to each modulator are shifted 90°, a single sideband transmission results without the use of expensive, sharp cutoff RF filters. The same holds true for AM stereo signal generation. If a small amount of phase shift is introduced into the audio driving the phase modulated path of the transmitter with reference to the envelope audio path, the signal will take on an unsymmetrical sideband structure. If the antenna system exhibits an overall tilt towards the upper sidebands, for example, one would only need to predistort the signal so the lower sidebands were favored in the transmitter, thereby restoring symmetry in the radiated signal.

The exciter can perform this task by inserting the delay sections into the left and right channels prior to matrixing. It is clear from Figure 4.2-A1 that this is the case for most C-QUAM exciters. The audio is delayed independently in the left and right channels. It is then matrixed to produce L + R and L − R that feed the QUAM generator. An uncompensated summation of L + R is used to feed the external broadcast transmitter. Such equalization, known as *differential equalization*, can be quite powerful in correcting situations where asymmetrical antenna sideband radiation would otherwise limit stereo performance.

The audio equalizer is broken into several components. The first is the *constant time delay circuitry* that is used to compensate for the bulk of differential transit delays through the transmitter. The second section is the *high frequency equalizer* that is used to compensate for additional phase shift (nonlinear group delay) characteristics in the broadcast transmitter audio path. These non-linear delay characteristics are introduced by rolloff in the modulator stages, phase shift in PDM filters, and reflected antenna impedances. The final section is the *low frequency equalizer* which, not unlike tilt correction commonly found in audio processors and limiters, anticipates the phase shifting and amplitude rolloff action of modulation transformers, reactors and coupling capacitors in older, plate modulated transmitters. By introducing a similar rolloff and phase shift into the phase modulated audio path, separation may be maintained at frequencies below 50 Hz.

The last section of the exciter to be discussed is the *pilot generator*. The *pilot tone*, a 25 Hz sine wave audio component, is injected into the L − R modulator at a 5% modulation level in the L − R channel. The purpose of the pilot tone is to indicate the presence of stereophonic information. This is different from the pilot tone signal in the FM stereo system where the 19 kHz signal is used as a synchronizing term in the demodulator. Indeed, AM stereo receivers could be built without the pilot tone detector, however, the consumer has grown accustomed to seeing a stereo indicator on the radio, hence it is incorporated into the system.

C-QUAM DECODING/RECEIVING TECHNIQUES

There are over a dozen ways to decode a C-QUAM transmission. The most common approach is the *feedback decoder* technique (see Figure 4.2-A2). In this drawing, there are three detectors: envelope, in-phase and quadrature.

The envelope detector demodulates the monophonic, L + R information. It may be a simple diode detector, however, most stereo demodulator integrated circuits utilize a limiter/multiplier approach that offers superior performance. Distortion measurements in the 0.1%–0.3% region are commonly found at 99% negative modulation when this technique is used.

The in-phase and quadrature demodulators are identical in action to the balanced modulators found in the exciter. Each has a reference signal, either 0° for the I detector, or 90° for the Q detector, and is phase

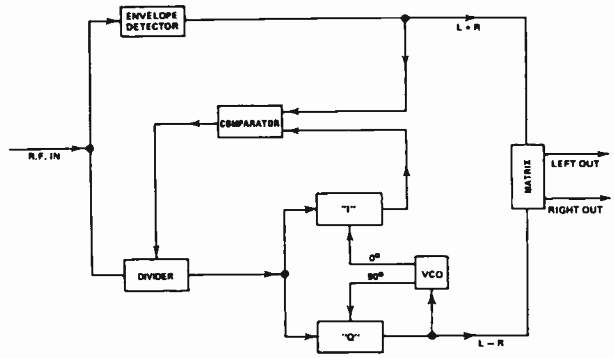


Figure 4.2-A2. Block diagram of feedback decoder.

locked to the incoming, received signal. However, referring back to the system equation, the actual received I signal is not 1 + L + R, but instead, (1 + L + R) cos ϕ, and the received quadrature signal is (L − R) cos ϕ rather than L − R as desired. In a proper C-QUAM decoder, the incoming RF/IF signal must be divided by the term cos ϕ, generally prior to quadrature demodulation. In the feedback decoder, the divider stage accomplishes this task. The derivation of the cos ϕ term conveyed in the phase modulated component is, however, an interesting process.

As indicated earlier, the I detector output would, since it is both phase and amplitude sensitive, be (1 + L + R) cos ϕ assuming no divider action. If the cos ϕ component is eliminated from the I detector output, the I detector would produce a signal identical in theory to the envelope detector. Assuming further that the I and Q detectors are fed from the same IF signal path, as cos ϕ is removed from the I detector output, it is also removed from the Q detector output, leaving L − R as the result. Since it is known that the IF signal must be divided by cos ϕ, a divider stage is placed in the IF path feeding the I and Q detectors. If the output of the I detector is then analyzed against the envelope detector output in a high gain comparator, the resultant is cos ϕ. By connecting the output of the comparator to the control port of the divider, a feedback loop results that will cause the cos ϕ signal normally found at the output of the I detector to be effectively transferred to the input of the divider. Since the IF path is therefore divided by cos ϕ, the Q detector performs the task of directly demodulating L − R.

The remaining circuitry in the decoder detects the pilot tone and dematrixes the L + R and L − R audio terms into the original left and right components.

PERFORMANCE CONSIDERATIONS

The C-QUAM system is capable of on-air performance figures in excess of 40 dB separation and under 1% distortion from 100–5000 Hz, particularly on newer solid state broadcast transmitters. It is not uncommon to obtain separation figures in excess of 30 dB from 50 Hz to the 10 kHz NRSC limit.

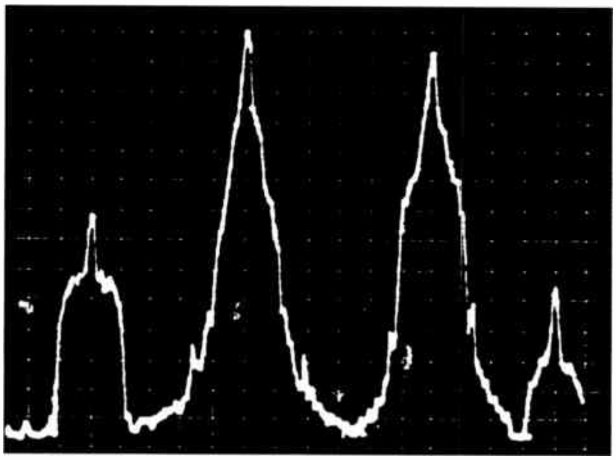


Figure 4.2-A3. Spectral photograph of four C-QUAM broadcasts.

RF spectra is also well controlled within the NRSC limits of occupied bandwidth now adopted by the FCC. A spectral photograph of four C-QUAM stations, three of which operate at a power level of 50 kW, depicts the spectral signal of C-QUAM broadcasts. It is virtually indistinguishable from monophonic broadcasts (see Figure 4.2-A3).

The broadcaster is free to choose the type of audio limiting desired for use with the C-QUAM signal with the exception that, if matrix type limiters are utilized, they must contain single channel limiting circuitry or the single channel limiter found in all C-QUAM exciters must be enabled. Since the L + R monophonic signal is summed and directly coupled to the broadcast transmitter, no additional tilt or overshoot is placed upon the audio allowing full modulation of the envelope component. The modulation constraints of the system are as follows:

- L + R (Monaural modulation) +125, −100% (FCC limits − higher mod is possible)
- L − R (stereo Difference channel only) ±100% (±45° phase modulation)
- L, R only (single channel, ref. to L + R) +125, −75% (limit of 71.56° phase modulation)
- *Pilot Tone* (injected into L − R channel) 25 Hz ± 0.1Hz, 5% +1/−0% injection level, sine wave.

In addition, audio preemphasis may be used. Motorola recommends use of the NRSC modified 75 μsec standard, however, tests with up to 30 dB preemphasis yield operation within FCC occupied bandwidth limits. The C-QUAM system has been used at audio bandwidths up to 15 kHz with excellent results, however, the FCC Rules limit audio bandwidth and subsequent RF emissions to 10 kHz as defined by the NRSC guidelines.

Use of the C-QUAM system will not reduce the monophonic coverage of the broadcast station. Since the envelope is sent as a compatible signal (the amplitude modulation of the station continues as 1 + L + R), the stereo component should pass undetected in existing mono receivers and full modulation of the envelope may be contemplated. The stereo receiver will experience a slight S/N degradation due to the fact that a second channel of information has been opened to allow passing of the stereo information. The degradation is usually under 3 dB and is typically 1.5 dB, which is nearly imperceptible under normal stereo programming conditions. All C-QUAM receivers contain circuitry that restores the receiver to mono reception when conditions warrant, thereby assuring no loss of coverage to the listener.

Receiver technology continues to expand with the advent of newer generation C-QUAM decoders. Original decoder IC's demodulated the incoming signal but performed very little conditioning of the signal to extend the performance of the receiver. Later generation decoders contain circuits that allow new receivers to be designed with adaptive bandwidth and synchronous detection features that improve reception under weak signal conditions. The MC13022 decoder IC, (see Figure 4.2-A4(a) and (b)), contains such circuitry that allows receiver designers to produce radios that operate at NRSC bandwidths under stronger signal conditions and automatically variable bandwidths sliding to 3–4 kHz under weak incoming signal conditions. The merits of such a system are obvious. The listener can enjoy wideband AM reception when signal conditions warrant while the receiver will automatically reduce the bandwidth under poor conditions, thereby retaining the listener. In addition, NRSC deemphasis is easily implemented with the MC13022.

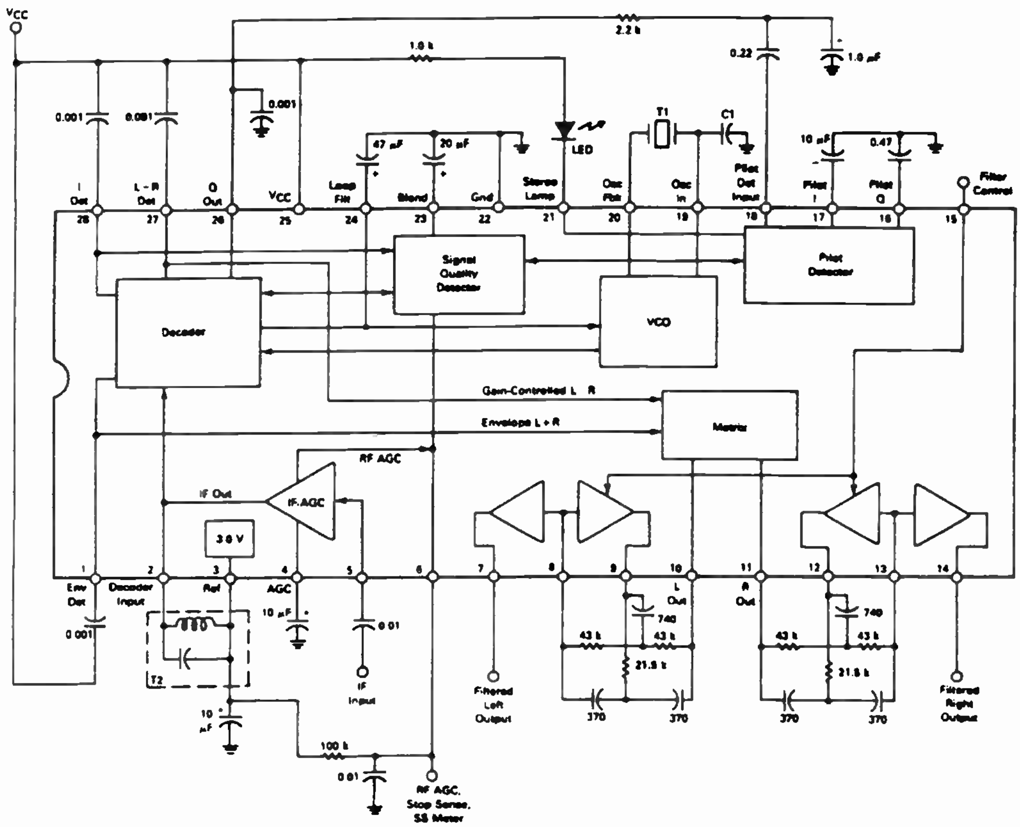


Figure 4.2-A4(a) Block diagram of the MC13022.

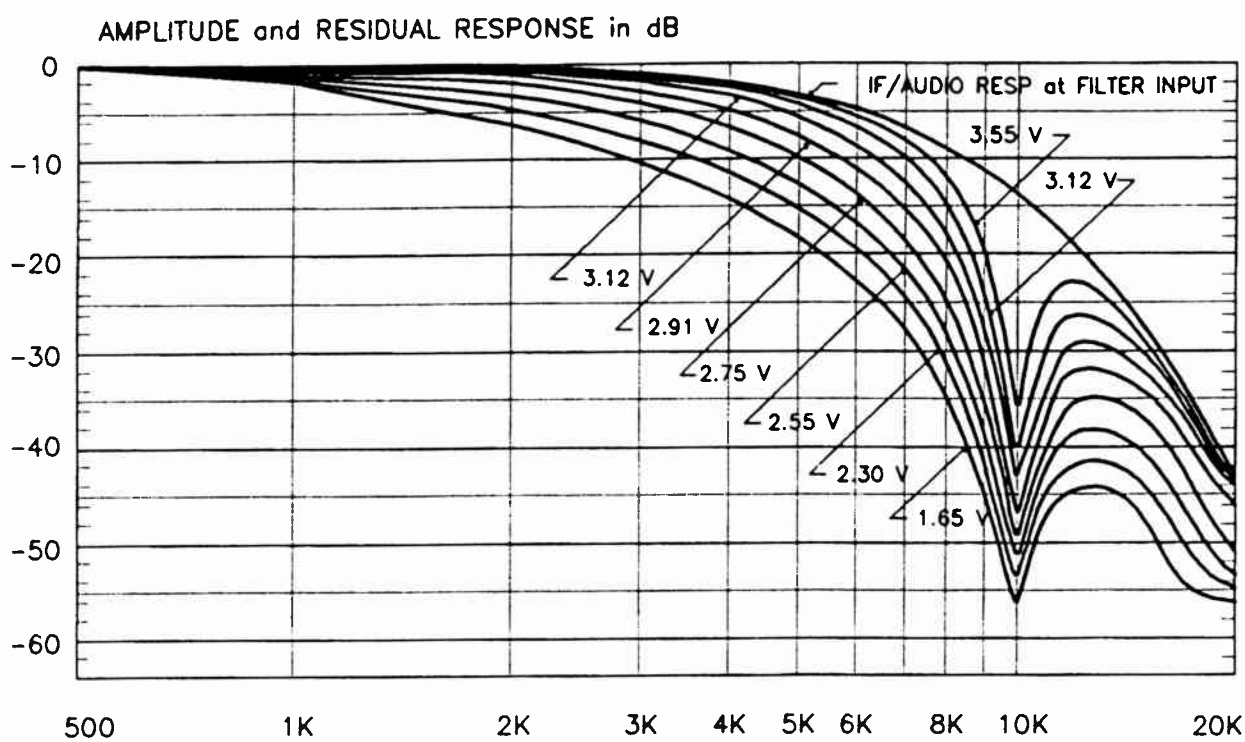


Figure 4.2-A4(b). MC13022 Variable Q notch filter response curves.

4.3 FM BROADCAST TRANSMITTERS

GEOFFREY N. MENDENHALL, P.E., RICHARD J. FRY, CPBE
HARRIS CORPORATION-BROADCAST SYSTEMS DIVISION, QUINCY, IL
WITH CONTRIBUTIONS BY: WARREN B. BRUENE, P.E., MUKUNDA B. SHRESTHA, AND EDWARD J. ANTHONY

INTRODUCTION

Although the mathematical principles explaining frequency modulation (FM) had been known for many years, the advantages and practical application to radio broadcasting were not realized until the 1930s, when Major Edwin H. Armstrong did extensive developmental work proving that FM radio transmissions were possible. Many theoreticians claimed to have proof that Armstrong's experiments were impossible based on mathematical models claiming that an infinite transmission bandwidth would be required. He never received proper credit for his many contributions to the radio communications industry during his lifetime.[1]

Among the advantages of FM are freedom from static, wide audio bandwidth and the ability of an FM receiver to capture the stronger of two signals transmitted on the same carrier frequency.

In 1940, following extensive public hearings, the Federal Communications Commission (FCC) established the FM Broadcast Service and set aside 40 channels in the 42 to 50 MHz band with commercial operation scheduled to begin January 1, 1941. Although World War II stopped all non-military radio construction, more than 40 FM stations continued to serve over 400,000 receivers. To eliminate the interference problems resulting from skywave reflection in the prewar FM band, the Commission moved the FM Broadcast Service to the 88 to 108 MHz band in 1945, thereby increasing the number of available channels to 100. However, the expected growth of FM broadcasting did not materialize. Since conversion of prewar FM receivers to the new band was not practical, purchase of a new receiver was the only way to receive the new FM stations. Television appeared to offer much more to the consumer than FM radio, since most FM stations merely duplicated the programming of an affiliated AM station.

Despite the potential for a higher quality broadcast service, there was little public demand for new FM receivers and virtually no public reaction when FM stations dropped popular programs from their schedule or were off the air due to equipment failures. It is not surprising, therefore, that in May 1950, there had been only 16 new FM license applications during the previous 15 months in which 259 FM stations ceased operations. It was not until after the introduction of stereo multiplex FM broadcasting that public awareness of the FM band increased and then skyrocketed in the 1970s to make FM the dominant medium for musical programming.

CHARACTERISTICS OF FM COMPARED TO AM

The 88–108 MHz FM broadcast band is relatively free of atmospheric and other noise interference. Emission at these frequencies is not propagated great distances by the ionosphere as it is in the 550–1600 kHz AM broadcast band. Therefore, noise from lightning discharges is limited to line-of-sight distances and is almost negligible. Man-made noise is a far greater source of interference, particularly in urban areas. The level of manmade noise falls off at increasing higher frequencies so that the microvolts-per-meter noise level is about one-tenth as great in the FM band as it is in the AM band.

In addition, FM has an improved noise threshold characteristic when compared to AM. Limiting circuitry symmetrically clips the RF waveform in an FM receiver to remove any amplitude variations produced by static or impulse noise before they reach the demodulator which responds only to phase or frequency changes in the signal. The FM improvement factor is illustrated in Figure 4.3-1. Note the sharp knee in the threshold curve above which noise and interference are suppressed, which results in an improved signal to noise (S/N) ratio. This same *capture* effect causes a weaker FM signal on the same channel to be suppressed, resulting in greatly reduced co-channel interference.

For even greater noise reduction, preemphasis (75 μsec time constant) is employed in the transmitter whereby the audio frequency components above about 2.1 kHz are boosted in amplitude at the rate of 6 dB per octave before being applied to the modulator. Flat frequency response is restored in the receiver's deemphasis network by attenuating the higher frequencies the same amount they were boosted in the transmitter. At the same time, the high frequency noise components which are characteristic to FM are also attenuated, resulting in greatly reduced background noise. Preemphasis is discussed in more detail in the FM theory and FM exciter sections of this chapter.

The net result of the above factors is a much better signal-to-noise ratio on FM than on AM. The lower background noise, along with the wider frequency response of FM versus AM (by virtue of its 200 kHz

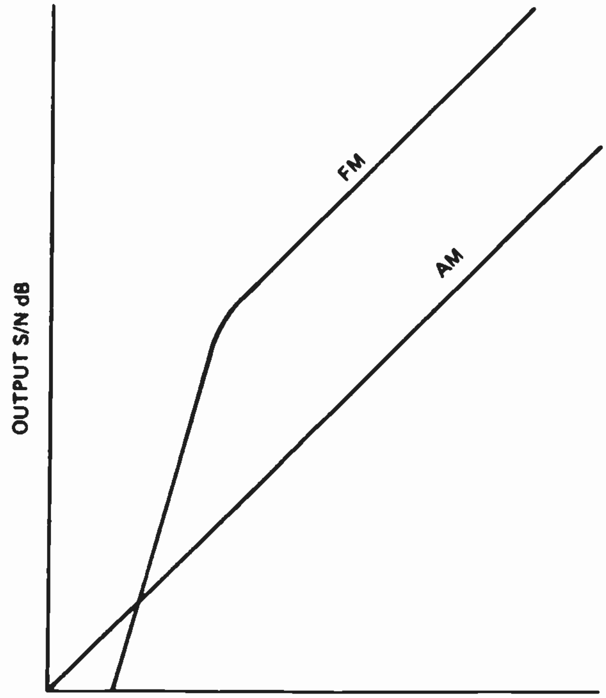


Figure 4.3-1. FM improvement factor.

channel bandwidth) means that a signal of higher quality and wider dynamic range can be enjoyed.

Occupied Bandwidth

The advantages of the FM broadcast medium over the AM broadcast medium do not come free of compromise. A standard FM broadcast channel occupies more than 10 times the bandwidth of an AM broadcast channel. This is because the more complicated nonlinear sideband structure of a frequency modulated carrier with wide deviation requires much more bandwidth than the simpler linear distribution of sidebands in an AM system. More details on the amplitude and spacing of FM sidebands are given in the section on FM modulation theory in this chapter.

High Fidelity

Uniform frequency response over the audible range of at least 50 Hz to 15 kHz, very low amplitude distortion (harmonic and intermodulation), very low noise level and good transient response (uniform time delay versus frequency) are necessary for hi-fi performance. The FM channel authorizations provide for adequate audio frequency response and a low-noise radio link to the listener. The rest of the performance is a matter of equipment design.

Stereophonic Transmission

The wide channel allocations and ability of FM to multiplex compatibly several audio channels on one carrier permitted development of a practical stereo broadcasting system. This has enabled the broadcast industry to provide the public with reproduction quality as good as or better than that which is available on stereo records or tapes. The advent of the compact disc and other digital audio source equipment will continue to challenge equipment manufacturers and station engineers further to improve the performance of the entire FM broadcast chain and to convert to an all digital audio transmission path. Recently introduced digital audio signal processing, uncompressed digital studio to transmitter links and digital DDS FM exciters now make it possible to 100% digitally convey the exact, "bit-for-bit," audio data from a compact disc to the FM transmitter power amplifier without any intervening analog to digital (A/D) conversions or distortions. Detailed technical information about stereophonic transmission theory and standards is provided in Chapter 4.4 "FM Stereo and SCA." See [27] at the end of this chapter for further information about digital FM exciter technology.

Subsidiary Communications Authorization (SCA)

The wide channel bandwidth authorized for FM broadcasting also makes it feasible to multiplex subsidiary audio or data channels commonly referred to by the FCC as Subsidiary Communications Authorization (SCA) together with the stereo transmission. SCA channels provide an important source of revenue to many stations as well as provide many useful audio and digital services to the community. Detailed technical information about SCA transmission theory and standards is provided in Chapter 4.4 "FM Stereo and SCA."

The AES/EBU Digital Serial Audio Data Interface Standard

Just as serial data can be used to pass information between computers, printers and modems, so various interconnect standards have been developed to transfer digitized audio data between components in an all-digital broadcast facility. AES3 is the digital audio data interface standard for audio sources, mixing equipment, audio processing equipment, studio to transmitter link (STL) equipment and transmitter inputs. The European Broadcasting Union (EBU) has republished a standard which is identical to the AES3 standards, except for the use of transformer-coupled inputs and outputs. The format of the AES3 data subframe is shown in Figure 4.3-2.

Some highlights of the AES3, 1992 (r1997) data format are:

- The interface format can accommodate 16, 20 or 24-bits of digital audio information
- The interface handles serial data transmission of two channels of digitized audio over distances of up to 500 ft using balanced cable specifically designed for the AES3 transmission standard. Transmission distances of up to 50 ft can be realized over a conventional shielded, twisted-pair, audio wire, when properly terminated

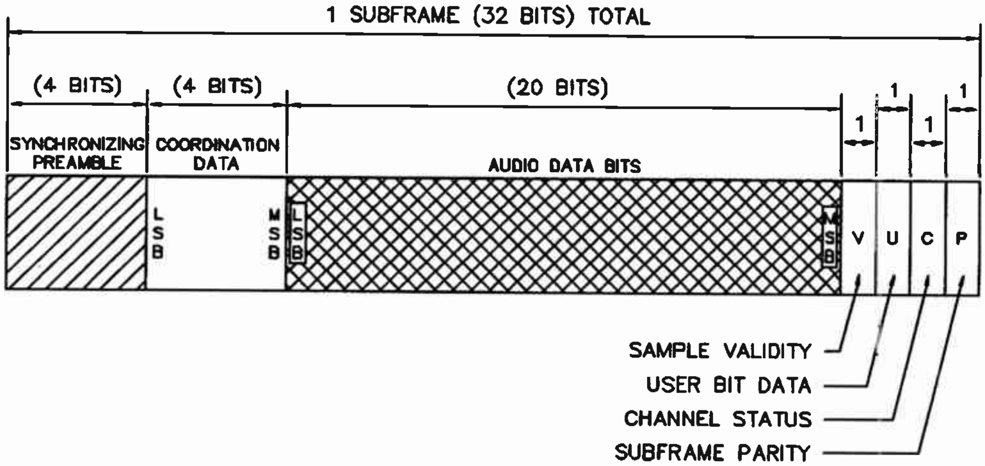


Figure 4.3-2. Thirty two-bit subframe format of AES/EBU serial data.

- The interface uses standard 3 pin, XLR-type connectors, carrying balanced, RS-422 compatible signals that are polarity independent. The input and output impedance for the interface is 110 ohms
- The data is sent least significant bit (LSB) first, with alternating subframes for Channel 1 and Channel 2
- The data is self-clocking and does not require an additional CLOCK connection to synchronize the source and destination. The SYNC data allows digital equipment to recognize the start of each 32-bit block of audio data and synchronize the master clocks
- There are four ancillary bits for validity, user bit, channel status and parity
- The validity bit indicates whether the audio sample data bits are valid and error free
- The user bit carries hardware or system specific information
- The channel status carries data concerning emphasis, sampling frequency and various other information
- The parity bit provides even parity over the current subframe, allowing simple detection of transmission errors.

The acceptance of the AES3 data interface standard by major broadcast equipment manufacturers makes it possible to build an all digital studio with an all digital link to the transmitter using standard off-the-shelf equipment. The standard AES3 transmit and receive chipsets support the three commonly used sampling rates of 48.0 kHz (studio), 44.1 kHz (consumer CD) and 32.0 kHz (broadcast transmission). Since the current FM stereo transmission standard limits the frequency response of the left and right channels to a maximum of 15 kHz, a 32.0 kHz sampling rate is often used. There are situations, particularly in the studio environment, where the higher 48.0 kHz sampling rate may be desired, so make sure that all the interfacing equipment has the ability to accept both the studio and broadcast transmission data rates. For further information on the AES3 standard and the transport of digital audio, see [28, 29, and 31].

FCC Transmission Standards

The FCC regulates and enforces the technical standards that apply to radio broadcasting in the United States. In theory, this ensures that the public is provided with a consistently high standard of transmission quality from station to station. The rules and regulations covering radio broadcast services, including those for FM broadcast transmitters are set forth in Part 73 of the Code of Federal Regulations (CFR) available from the U.S. Government Printing Office in Washington, D.C. The rules and regulations are changed from time to time to keep pace with new technology and changes within the broadcast industry. Every broadcast engineer should have access to a current copy these rules and regulations so that the station's technical performance is maintained within the prescribed limits.

FREQUENCY MODULATION THEORY

Angular Modulation

Frequency modulation (FM) and phase modulation (PM) are both special cases of angular modulation. In any angular modulation system both the frequency and phase of the carrier vary with time as a function of the modulating signal.

The relationship between the frequency deviation of the carrier, the phase deviation of the carrier and the sinusoidal modulating frequency is defined as the modulation index (m).

Where:

$$m = \frac{\text{frequency deviation (peak-to-peak Hertz)}}{\text{modulating frequency (Hertz)}} \quad (1)$$

Since frequency modulation and phase modulation are both subsets of angular modulation, they are virtually indistinguishable from one another except in the modulator characteristics.

In a PM system, the modulating signal causes the phase of the carrier wave to vary according to the instantaneous amplitude of the modulating signal. A phase modulator generates a constant amount of phase deviation of the carrier with a constant amplitude modulating signal independent of the frequency of the modulating signal. The frequency deviation of the carrier produced by a phase modulator does increase as the modulating frequency is increased even though the level of the modulating voltage is held constant. The net effect is that the phase modulator behaves as if it were a frequency modulator with a 6 dB/octave rising slope on the modulating signal input.

A frequency modulator generates a constant frequency deviation of the carrier with a constant amplitude modulating signal independent of the frequency of the modulating signal. The phase deviation of the carrier produced by a frequency modulator decreases as the modulating frequency is increased even the though the level of the modulating voltage is held constant. The net effect is that the frequency modulator behaves as if it were a phase modulator with a 6 dB/octave falling slope on the modulating signal input.

In FM broadcasting, the signal carrier should have frequency deviation that is proportional to the amplitude of the modulating signal, but independent of the frequency of the modulating signal as generated by a frequency modulator.

The instantaneous frequency (rate of change of phase) of the RF output wave differs from the carrier frequency by an amount proportional to the instantaneous amplitude of the modulating waveform. For example, consider a 100 MHz carrier wave frequency modulated by a 1,000 Hz audio tone and assume that a 1 volt input to the modulator causes ±20 kHz of frequency deviation on the positive and negative peaks of this tone. If the audio input amplitude is increased to 2 V, the peak deviation will become ±40 kHz varying in sine-wave fashion from one peak of deviation to the other and back again at the 1,000 Hz rate. In FM broadcasting, the FCC has defined 100% modulation as modulation which results in a peak frequency deviation of ±75 kHz of the RF carrier.

When preemphasis is used ahead of the frequency modulator, the system becomes a phase modulator at audio frequencies above the turnover point of the preemphasis network. This is because the frequency response of the preemphasis network rises at the rate of 6 dB/octave above this point. FM broadcasting with preemphasis really becomes a mixture of FM at low modulating frequencies and PM at high modulating frequencies.

FM Sideband Structure

The frequency modulated RF output spectrum contains many sideband frequency components, theoretically an infinite number. Consider, as an example, an RF carrier of frequency f_c modulated by a sinusoidal signal with frequency f_m ($f_m < f_c$). The spectrum consists of pairs of sideband components spaced from the carrier frequency by multiples of the modulating frequency. When the modulation index is small (m = 0.5), the amplitude of the second and higher order sidebands is small so that the output consists mainly of the carrier and the pair of first-order sidebands, as illustrated in Figure 4.3-3(a). The total transmitter RF output power remains constant for a given modulation waveform, but the distribution of that power into the sidebands varies with the modulation index such that power at the carrier frequency (and lower order sidebands) is reduced by the amount of power added to the higher order sidebands.

As the modulation index is increased (as in wide deviation FM broadcasting), the higher order sidebands become more prominent. The amplitude and phase of the carrier as well as the sidebands can be expressed mathematically by making the modulation index (m) the argument of a simplified Bessel function:

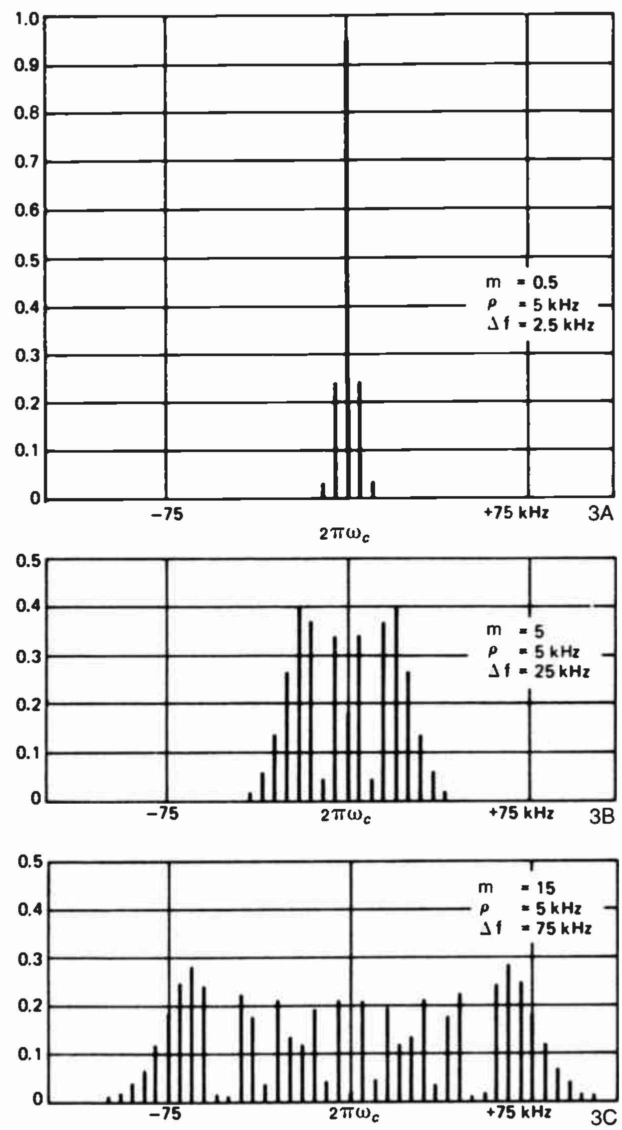


Figure 4.3-3(a), (b), (c). RF spectrum with modulation indexes of 0.5, 5.0 and 15.

$E(t)$ = Total RF Output Voltage (2)

$A[J_0(m) \sin \omega c(t)]$ carrier amplitude
$+[J_1(m) \sin (\omega c+\omega m)t]$ first order upper sideband
$-[J_1(m) \sin (\omega c-\omega m)t]$ first order lower sideband
$+[J_2(m) \sin (\omega c+2\omega m)t]$ second order upper sideband
$+[J_2(m) \sin (\omega c-2\omega m)t]$ second order lower sideband
$+[J_3(m) \sin (\omega c+3\omega m)t]$ third order upper sideband
$-[J_3(m) \sin (\omega c-3\omega m)t]$ third order lower sideband
$\pm[J_n(m) \sin (\omega c\pm n\omega m)t]$ higher order sidebands

where:

A = Unmodulated carrier amplitude constant
$J_0(m)$ = (modulated) carrier amplitude (a function of modulation index)
$J_1(m)$ = Amplitude of first order sidebands
$J_2(m)$ = Amplitude of second order sidebands
$J_n(m)$ = Amplitude of nth order sidebands
m = The modulation index
ωc = $2\pi F_c$ (The carrier frequency in radians/second)
ωm = $2\pi F_m$ (The modulating frequency in radians/second)

The numeric values of the Bessel functions, $J_0(m)$ through $J_n(m)$, which express the amplitudes of the various frequency components can be found in mathematical tables. Figure 4.3-4 shows a graphical representation of how the Bessel function values for the carrier and the first eight pairs of sidebands vary with the modulation index.

In a monophonic FM broadcast transmitter, the modulation index can become very high at low modulating frequencies. With a 50 Hz audio input signal of sufficient amplitude to produce 75 kHz deviation (100% modulation), the modulation index is:

$$m = \frac{75{,}000}{50} = 1{,}500 \qquad (3)$$

With a 15,000 Hz input at the same deviation (also 100% modulation), the modulation index is only:

$$m = \frac{75{,}000}{15{,}000} = 5 \qquad (4)$$

Figures 4.3-3(b) and 4.3-3(c) illustrate the frequency components present for modulation indices of 5 and 15. Note that the number of significant sideband components becomes very large with a high modulation index. For a given modulation index, the total bandwidth occupied can extend beyond ±75 kHz from the carrier depending upon the modulating frequency.

This single tone modulating frequency analysis is useful in understanding the general nature of FM and for making tests and measurements. When audio program modulation is applied, there are many more sideband components present. They vary so much that sideband energy becomes distributed over the entire occupied bandwidth rather than appearing at discrete frequencies.

Bessel Nulls

At certain modulation indices, the amplitude of the carrier component of the signal goes to zero, with all the transmitted power being distributed at frequencies other than the carrier frequency. This carrier null phenomenon is useful as an extremely accurate method for measuring the frequency deviation and to check the calibration of modulation monitors. Referring again to Figure 4.3-4, note that the carrier amplitude goes to zero and reverses sign at several values of modulation index including: 2.405, 5.520 and 8.654. Figure 4.3-5

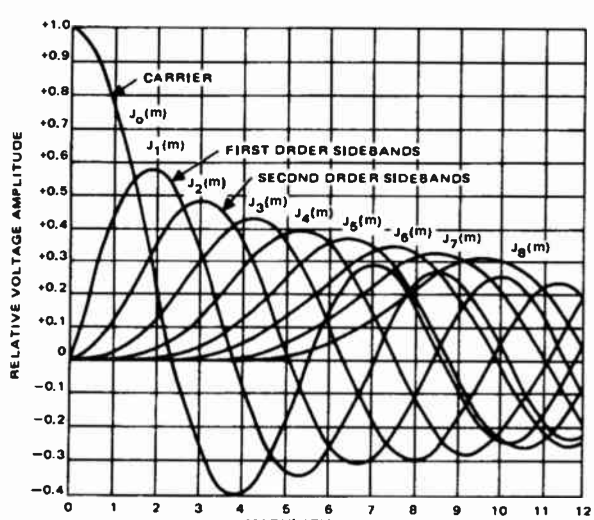


Figure 4.3-4. Relationship of carrier and sideband amplitudes to modulation index.

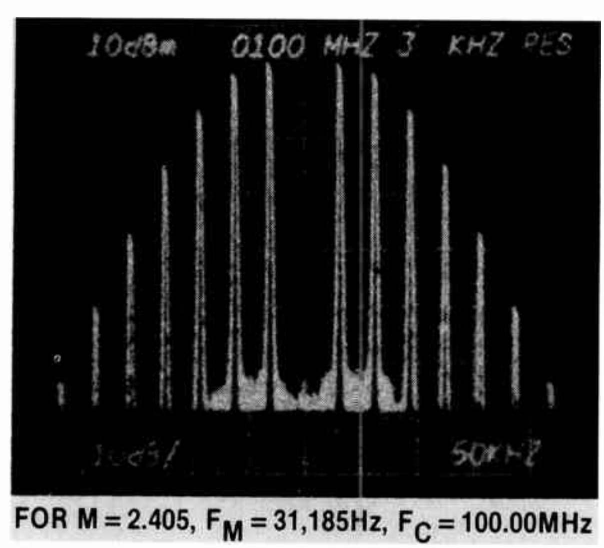


Figure 4.3-5. FM RF spectrum showing occupied bandwidth and Bessel Null of a single tone at ±75 kHz deviation.

Table 4.3-1
Sideband nulls as a function of modulation index and modulating frequency.

	Modulation Index (M)		(F_m) FOR 75 kHz Deviation	
Null	Carrier	First Sidebands	Carrier	First Sidebands
1ST	2.405	3.832	31,187	19,574
2ND	5.520	7.016	13,587*	10,690
3RD	8.654	10.174	8,667	7,372
4TH	11.792	13.324	6,361	5,629
5TH	14.943	16.471	5,023	4,554
6TH	18.071	19.616	4,150	3,823
7TH	21.212	22.759	3,536	3,295

* This tone is recorded on track 38 of the NAB Broadcast Audio System Test CD.

is a photograph taken from an RF spectrum analyzer showing the first Bessel null ($M = 2.405$) of a carrier at a frequency of 100 MHz.

To determine the audio input level required to achieve 75 kHz deviation, apply an audio tone of exactly 8,667 Hz (75,000 divided by 8.654). Starting from zero amplitude, increase the audio level until the carrier disappears for the third time (since 8.654 corresponds to the third Bessel null). At this audio level, the deviation is exactly 75 kHz. The carrier amplitude (null) detector must have sufficient selectivity to separate the carrier from the sidebands and could be a spectrum analyzer or a receiver with a narrow IF bandwidth. The FM signal can be heterodyned down to a convenient frequency for measurement, since heterodyning does not alter the modulation index. However, when a frequency (or phase) modulated wave is multiplied or divided, this also multiplies or divides the frequency deviation and the modulation index by the same amount.

A listing of useful carrier and first order sideband nulls as function of the modulation index (M) and the modulating frequency (F_m) is given in Table 4.3-1.

Occupied Bandwidth

After examining the Bessel functions and the resulting spectra, it becomes clear that the occupied band-width of an FM signal can be far greater than the frequency deviation of the signal. In fact, the occupied bandwidth is infinite (if all the sidebands are taken into account), so it is now clear that a frequency modulation system would require the transmission of an infinite number of sidebands for perfect demodulation of information. In practice, a signal of acceptable quality can

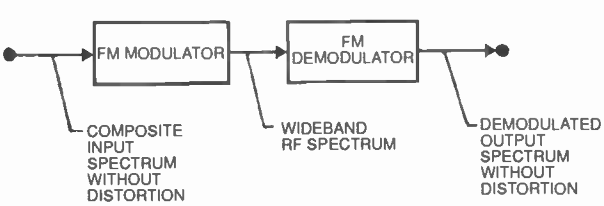
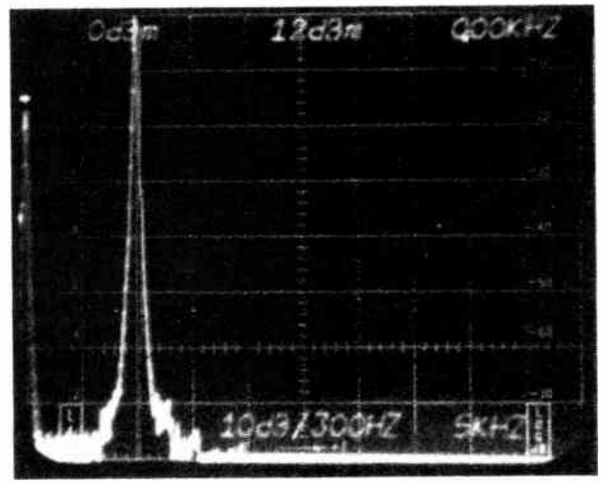
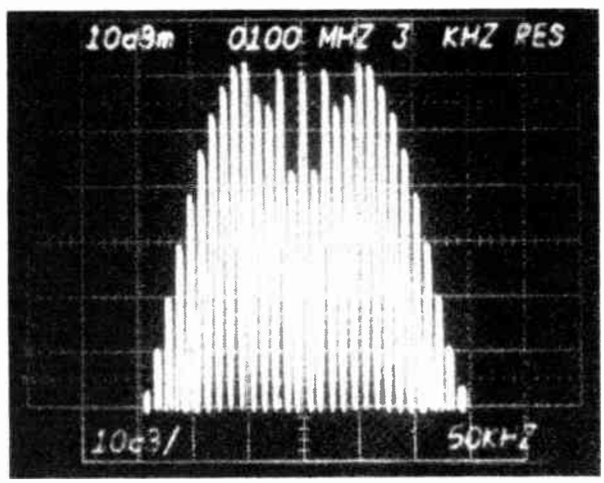


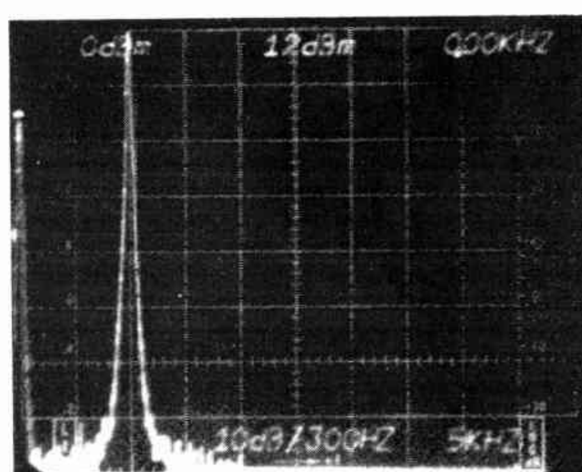
Figure 4.3-6(a). Wideband RF path.



BASEBAND SPECTRUM TO FM MODULATOR



RF SPECTRUM TO DEMODULATOR



DEMODULATED BASEBAND SPECTRUM

Figure 4.3-6(b). Single tone (10 kHz) modulation through wideband RF path.

be transmitted in the limited bandwidth assigned to an FM channel.

Effects of Bandwidth Limitation

Practical considerations in the transmitter RF circuitry make it necessary to restrict the RF bandwidth to less than infinity. As a result, the higher order sidebands will be altered in amplitude and group delay (time). Bandwidth limitation will cause distortion in any FM system.

Consider the block diagram shown in Figure 4.3-6(a), where a perfect FM modulator is connected to a perfect demodulator via an RF path of infinite bandwidth. The demodulated audio shown in Figure 4.3-6(b) contains no distortion components.

In Figure 4.3-7(a), a bandpass filter is inserted between the modulator and demodulator in order to restrict the bandwidth. Audio distortion products now appear at the output of our perfect demodulator as shown in Figure 4.3-7(b). These distortion products are due solely to the bandwidth restriction (300 kHz BW_3) imposed by the bandpass filter.

Figures 4.3-8(a) and 4.3-8(b) show the effects of a narrowband RF bandpass filter on the RF spectrum of a composite signal consisting of a stereophonic subcarrier modulated only on the left channel with 4.5 kHz plus a 67 kHz unmodulated SCA subcarrier. In Figure 4.3-8(a) through a wideband RF path, there are no baseband distortion products in the demodulated spectrum. Figure 4.3-8(b) shows the corresponding effects observed on the demodulated baseband spectrum for the same signal through a narrowband filter. The only distortion evident on the RF spectrogram is the loss of some sidebands greater than 150 kHz from the center frequency and some amplitude differences between the upper and lower sideband pairs. Note the creation of many undesired intermodulation terms in the demodulated baseband spectrum which cause crosstalk into both the stereophonic and SCA subcarrier bands. The change in the RF spectrum is subtle, but the resulting spectrum after demodulation is clearly modified.

As one can see, the distortion in any practical FM system will depend on the amount of bandwidth available as well as the modulation index being transmitted.

Group Delay Symmetry versus Amplitude Response Symmetry

While both the amplitude response and time response (group delay) have an effect on the amount of distortion added to the FM signal, the symmetry of the group delay response is more important than the total group delay variation or the amplitude response.

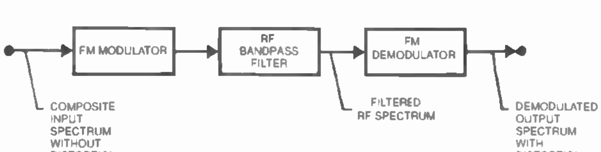
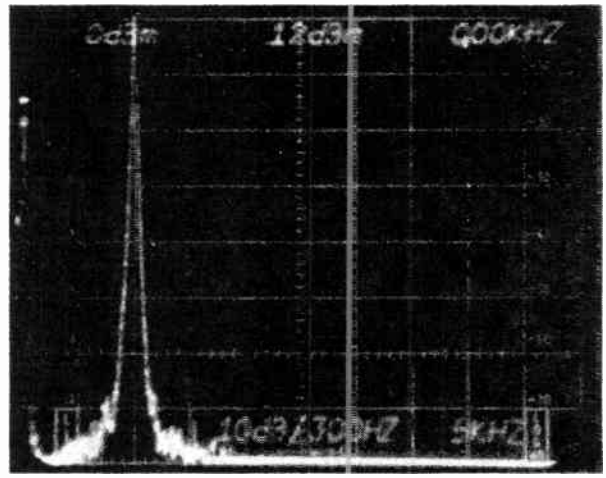
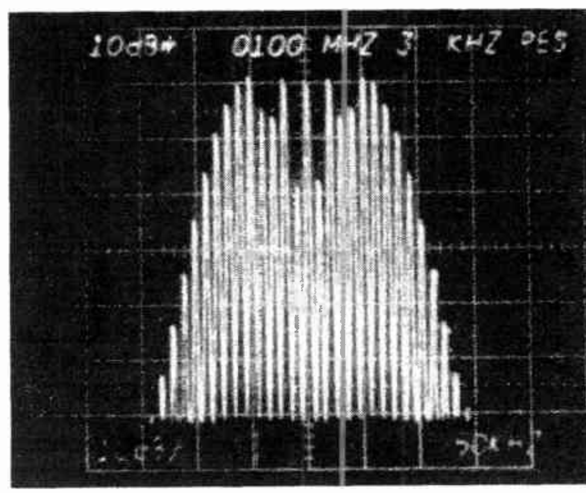


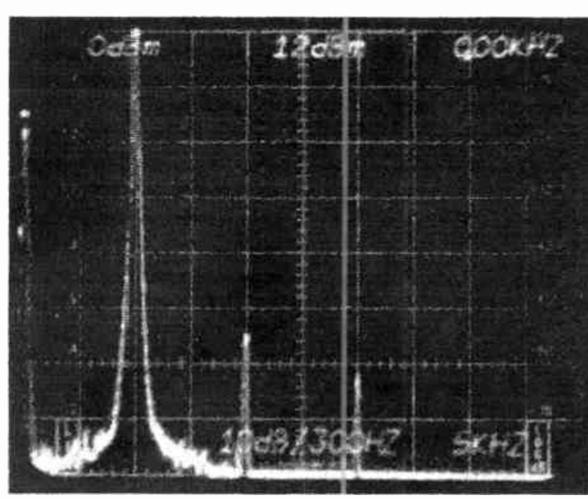
Figure 4.3-7(a). Bandwidth limited RF path.



BASEBAND SPECTRUM TO FM MODULATOR

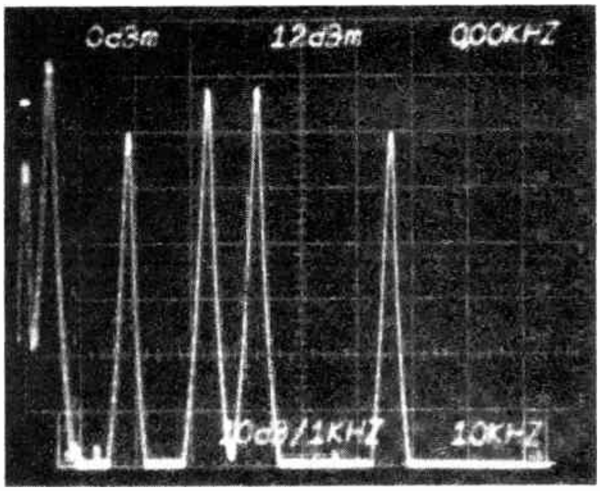


BANDWIDTH LIMITED RF SPECTRUM TO DEMODULATOR

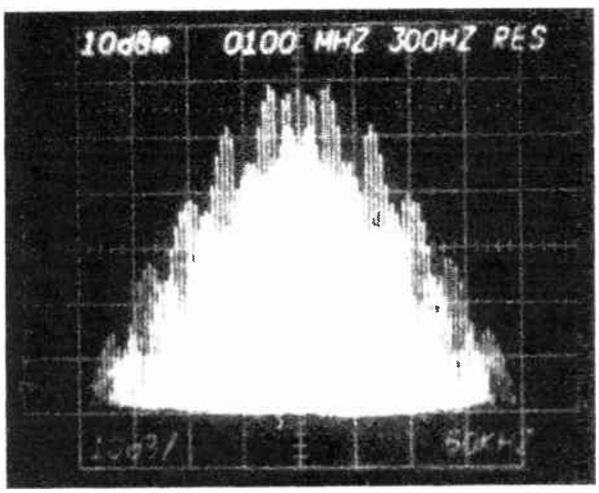


DEMODULATED BASEBAND SPECTRUM

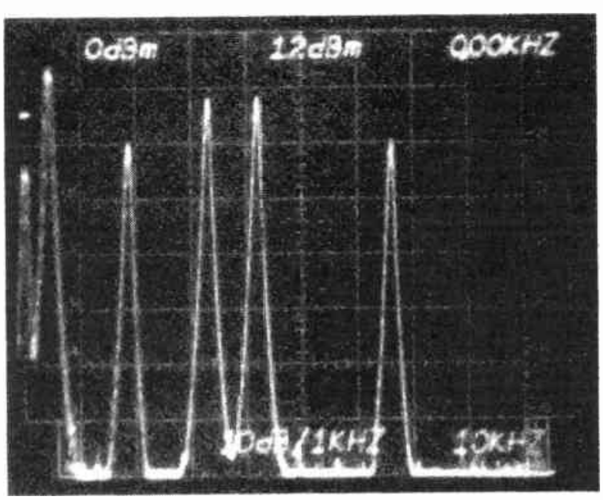
Figure 4.3-7(b). Single tone (10 kHz) modulation through narrow band RF path.



BASEBAND SPECTRUM TO FM MODULATOR

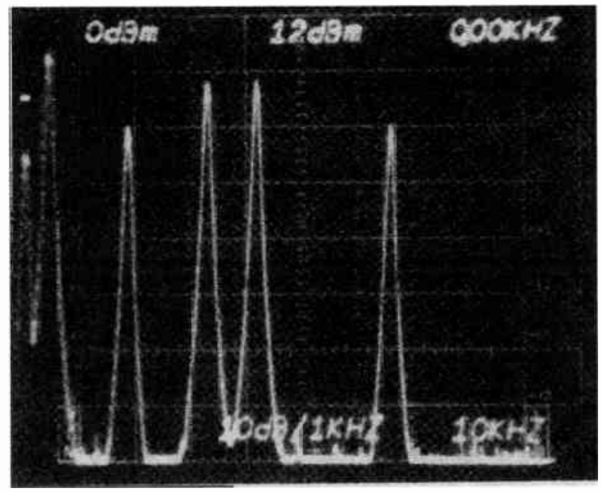


RF SPECTRUM TO DEMODULATOR

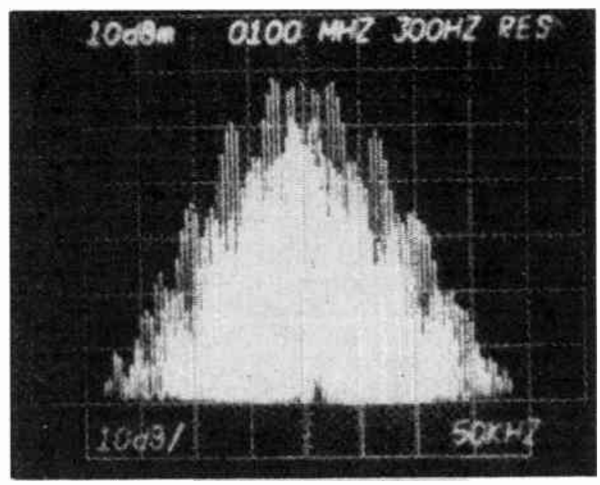


DEMODULATED BASEBAND SPECTRUM

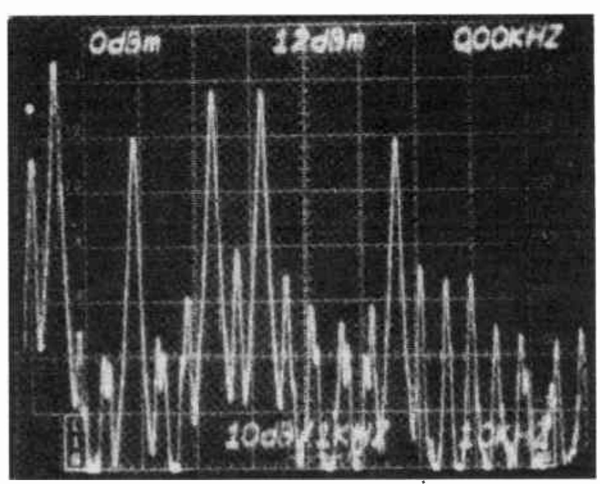
Figure 4.3-8(a). Stereo (L or R = 4.5 kHz) plus SCA (unmcd) modulation through wideband RF path.



BASEBAND SPECTRUM TO FM MODULATOR



BANDWIDTH LIMITED RF SPECTRUM
TO SPECTRUM



DEMODULATED BASEBAND SPECTRUM

Figure 4.3-8(b). Stereo (L or R = 4.5 kHz) plus SCA (unmod) modulation through narrow band RF path.

Best FM modulation performance is always obtained when the system is tuned for symmetrical group delay (time) response. Depending on the circuit topology, the tuning conditions for symmetrical group delay response may not coincide with the symmetrical amplitude response.

Limiting Factors within an FM Transmitter

Relating the specific quantitative effect of the bandwidth limitations imposed by a particular transmitter to the actual distortion of the demodulated composite baseband is a complicated problem indeed. Some of the factors involved are:

- Total number of tuned circuits involved
- Amplitude and group delay response of the total combination of tuned circuits in the RF path
- Amount of drive (saturation effects) to each Class C stage
- Nonlinear transfer function within each amplifier stage.

Improvement of the RF Path

The following design techniques can help improve the transmitter's bandwidth:

- Maximize bandwidth by using a broadband exciter and a broadband intermediate power amplifier (IPA) stage
- Use a single-tube design or a broadband, completely solid-state design where feasible
- Optimize both grid circuit and plate circuit of the tuned stage for the best possible bandwidth and symmetrical group delay response
- Minimize the number of interactive tuned networks
- Use a broadband antenna system that that provides a low standing wave ratio on the transmission line. For more detailed information about FM and digital modulation theory, see [2, 3, 4, 27, and 32].

PREEMPHASIS

The standards adopted for FM broadcasting in the U.S. require the use of preemphasis. The standard preemphasis curve is defined as an ideal RC network with a time constant of 75 microseconds. The 3 dB point is at a frequency of:

$$f = \frac{1}{2\pi(RC)} = \frac{1}{2\pi(75 \times 10^{-6})} = 2{,}122 \text{ Hz} \quad (5)$$

The 75 microsecond curve and the tolerance allowed by the FCC are shown in Figure 4.3-9. The frequency response characteristics in decibels for several popular preemphasis time constants are given in Table 4.3-2.

The reduction in receiver output noise due to the use of preemphasis in monophonic transmission is illustrated in Figure 4.3-10 (for 75 μsec preemphasis case). The noise voltage in a narrow bandwidth (for example, 1 Hz) increases directly with frequency, therefore, the power spectral density increases as the square of frequency as shown. When preemphasis is used, the noise voltage is attenuated above 2.1 kHz so that it remains constant with frequency. The power spectral density is also constant above 2.1 kHz. The area between these curves represents the noise power that is removed by the use of preemphasis. This diagram indicates the importance of preemphasis for high

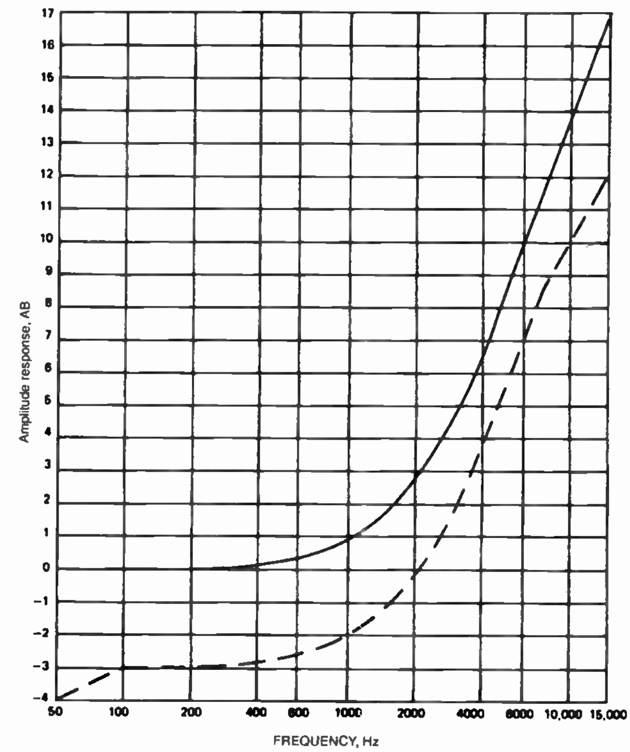


Figure 4.3-9. FCC Standard 75 microsecond preemphasis curve (solid line) and tolerance limits (solid and dashed lines).

Table 4.3-2
Precise amplitude response values for various amounts of preemphasis.

	Preemphasis Amplitude Response vs. Time Constant			
Freq (Hz)	25 μs	50 μs	75 μs	150 μs
50	0.000 dB	0.001 dB	0.002 dB	0.010 dB
100	0.001 dB	0.004 dB	0.010 dB	0.038 dB
400	0.017 dB	0.068 dB	0.152 dB	0.577 dB
1,000	0.106 dB	0.409 dB	0.871 dB	2.761 dB
2,000	0.409 dB	1.445 dB	2.761 dB	6.583 dB
3,000	0.871 dB	2.761 dB	4.769 dB	9.540 dB
4,000	1.445 dB	4.115 dB	6.583 dB	11.822 dB
5,000	2.087 dB	5.400 dB	8.164 dB	13.656 dB
6,000	2.761 dB	6.583 dB	9.540 dB	15.182 dB
7,000	3.442 dB	7.661 dB	10.749 dB	16.486 dB
8,000	4.115 dB	8.643 dB	11.822 dB	17.623 dB
9,000	4.769 dB	9.540 dB	12.785 dB	18.630 dB
10,000	5.400 dB	10.362 dB	13.656 dB	19.534 dB
11,000	6.005 dB	11.120 dB	14.451 dB	20.353 dB
12,000	6.583 dB	11.822 dB	15.182 dB	21.103 dB
13,000	7.135 dB	12.475 dB	15.858 dB	21.793 dB
14,000	7.661 dB	13.084 dB	16.486 dB	22.433 dB
15,000	8.164 dB	13.656 dB	17.073 dB	23.029 dB

Notes: 1) Values in shaded boxes subject to rolloff by audio low pass filter section of subcarrier generator. 2) Values shown in decibels relative to f_m = 0 Hz.

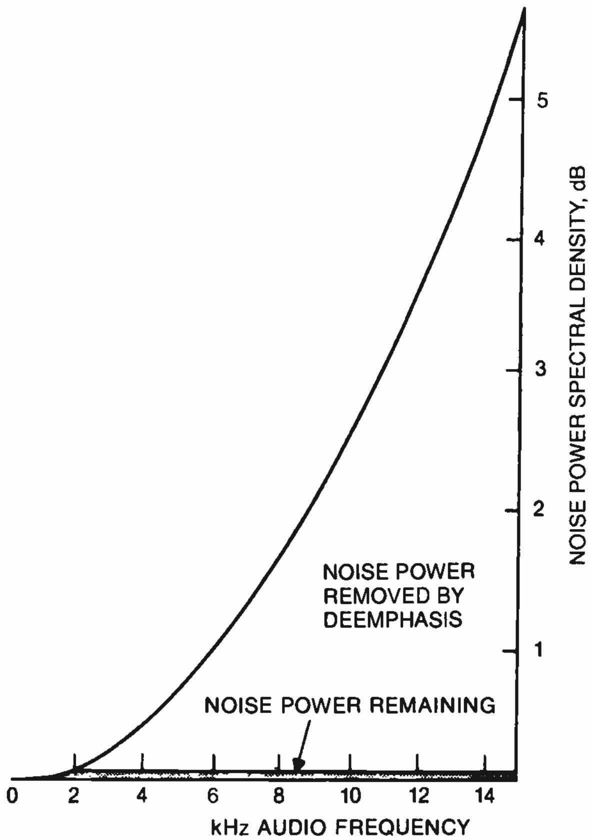


Figure 4.3-10. Noise power spectral density before and after deemphasis in receiver.

fidelity transmission because the high-frequency noise at the receiver would be very much greater without deemphasis.

Preemphasis is practical because program energy tends to peak at several kilohertz and then falls off fairly rapidly at the higher frequencies. For this reason, the higher frequencies can be boosted in amplitude without causing much increase in modulation level. There is some increase, however, so the net improvement due to preemphasis is the ratio of the areas under the two curves of the diagram less this reduction in audio input level required to keep within the 100% modulation limit. Modern audio processing equipment takes the preemphasis curve into account when controlling peak modulation levels.

In the section on FM modulation theory, it was mentioned that the use of preemphasis ahead of an FM modulator actually causes the system to behave like FM at low modulating frequencies and like PM at high modulating frequencies. The location of the preemphasis network in the system depends on whether the station is operating in the monaural or stereo mode. In the case of monaural transmission, the preemphasis network is usually located in the FM exciter just ahead of the modulator stage. Stereo transmission requires that the FM modulator have a flat response to the composite baseband signal from the stereo generator, so the individual preemphasis networks for the left and right channels are located in the stereo generator before the left and right audio channels are multiplexed into the composite baseband signal. In the case of a digital stereo generator, the AES3 serial audio data contains both the left and right channel information in alternating data frames. The digital stereo generator applies the preemphasis required and 15 kHz audio low pass filtering using Digital Signal Processing (DSP) techniques.

FM TRANSMITTER POWER OUTPUT REQUIREMENTS

The FCC regulates the power of FM broadcast stations in terms of effective radiated power (ERP) which is determined by the class of station and the antenna height above average terrain (HAAT). The authorized ERP applies only to the horizontally polarized component of radiation. Elliptical or circular polarization is also permitted, in which the ERP of the vertically polarized component may be as great as the authorized horizontal component. This means that twice as much total power may be radiated and twice as much transmitter power will be required.

The transmitter power requirement can be reduced by increasing the gain of the antenna. There is, of course, an economic trade off between the cost of a higher gain antenna versus the cost of a larger transmitter and the added primary power costs. For a high ERP, it is common to use antennas with up to 12 elements, providing a power gain of about 12.6 (or 6.3 in each polarization).

The long transmission lines associated with the tall towers commonly used are a source of considerable power loss. For example, the efficiency of 2,000 ft of 3⅛ in rigid coax at 100 MHz is only about 62%.

After the required ERP is determined, the transmitter power output (TPO) can be calculated taking into account the transmission line losses and the antenna array gain. Depending on the particular situation, the TPO could vary from as little as 50 W to as much as 70 kW.

FM transmitters are designed to operate over a wide range of power outputs so that with a few basic models any required power output can be furnished. Popular maximum ratings range from 250 W to 70 kW. Most installations use a maximum TPO of 30 kW or less because it is more economical to achieve the maximum 100 kW ERP with circular polarization by means of sufficient antenna gain.

THE FM TRANSMITTER

The purpose of the FM transmitter is to convert a main channel audio (mono or stereo) frequency signal and its associated audio or data subcarriers, or an AES3 serial digital audio data bitstream into a frequency modulated, radio frequency signal at the desired power output level to feed into the radiating antenna system. In its simplest form, it can be considered to be an FM modulator and an RF power amplifier packaged into one unit as shown in Figure 4.3-11.

Actually, the FM transmitter consists of a series of individual subsystems each having a specific function:

1. The FM exciter converts the analog audio baseband or serial digital audio data into frequency modulated RF and determines the key qualities of the signal
2. The IPA is required in some transmitters to boost the RF power level up to a level sufficient to drive the final stage
3. The final power amplifier further increases the signal level to the final value required to drive the antenna system
4. The transmitter control system monitors, protects and provides commands to each of these subsystems so that they work together to provide the desired result
5. The RF low pass filter removes undesired harmonic frequencies from the transmitter's output, leaving only the fundamental output frequency
6. The directional coupler provides an indication of the power being delivered to and reflected from the antenna system
7. The power supplies convert the input power from the ac line into the various dc or ac voltages and currents needed by each of these subsystems

FM Exciter

The heart of an FM broadcast transmitter is its exciter. The function of the exciter is to generate and modulate the carrier wave with one or more inputs (mono, stereo, SCA) in accordance with FCC standards. The FM modulated carrier is then amplified by a wideband amplifier to the level required by the transmitter's following stage.

Stereo transmission places the most stringent performance requirements upon the exciter. Since the exciter is the origin of the transmitter's signal, it determines most of the signal's technical characteristics including S/N, distortion, amplitude response, phase response and frequency stability. Waveform linearity, amplitude bandwidth and phase linearity must be maintained within acceptable limits throughout the analog baseband chain from the stereo and subcarrier generators to the analog FM exciter's modulated oscillator. The recent introduction of AES3 digital audio transport and all digital FM modulation techniques like DDS eliminate the distortions introduced by analog circuits. In a digital FM exciter, the left and right audio data is converted into a digital representation of stereo baseband by DSP. This data is then further converted into a frequency modulated carrier by a DDS numerically controlled oscillator (NCO). From here, the FM carrier is usually amplified in a series of Class C nonlinear power amplifiers, where any amplitude variation is removed. The amplitude and phase responses of all the RF networks which follow the exciter must also be controlled to minimize degradation of the signal quality.

Indirect FM

Before the advent of stereo broadcasting, most of the FM exciters employed phase modulation techniques. Some of these were adapted to stereo but it was difficult to achieve and maintain the performance requirements for stereo transmission.

In the FM modulation theory section of this chapter, the important relationship between PM and FM was discussed. If the audio frequency response is made to fall off at the rate of 6 dB/octave across the entire audio band at the input of a phase modulator, the resulting modulated output will be identical to that of a frequency modulated carrier. In 1948, James R. Day, an associate of Edwin H. Armstrong, demonstrated a Serrasoid phase modulator that used pulse circuits to achieve large phase shifts with low distortion. The Serrasoid modulation technique, in conjunction with the audio response shaping mentioned above, became the standard method of generating wide deviation FM with low distortion.

Known as indirect FM, this principle was used in many FM broadcast transmitters prior to the advent of stereo broadcasting. The advantage was that the carrier frequency could be generated by a stable crystal oscillator with the phase modulation occurring in later stages. The amount of phase deviation with low distortion was limited in most systems, so it was necessary to start with a low frequency oscillator and multiply its frequency and modulation index many times to achieve 75 kHz deviation at low modulating frequencies. This technique has been abandoned in favor of direct FM systems.

Direct FM

Direct FM is a modulation technique where the frequency of an oscillator can be made to change in proportion to an applied voltage. Such an oscillator, called a voltage tuned oscillator (VTO), was made possible by the development of varactor tuning diodes which change capacitance as their reverse bias voltage is varied (also known as a voltage controlled oscillator or VCO).

If the composite baseband signal is applied to the tuning terminal of a VTO, the result is a direct FM modulated oscillator. Figure 4.3-12 is a block diagram that fits most of the modern direct FM exciters on the market.

The S/N of an FM exciter is dependent on the short-term stability of the modulated oscillator. That stability is determined by factors such as operating level, noise figure of the oscillator transistor, circuit configuration, method of amplitude limiting, loaded "Q" of the oscillator tank circuit and the mechanical stability of components. Optimization of these factors has resulted in a S/N of better than 90 dB below 100% modulation in the current generation of analog FM exciters.

Analog FM Modulator Linearity

Nonlinearities in the FM oscillator can, by altering the waveform of the baseband signal, create distortion

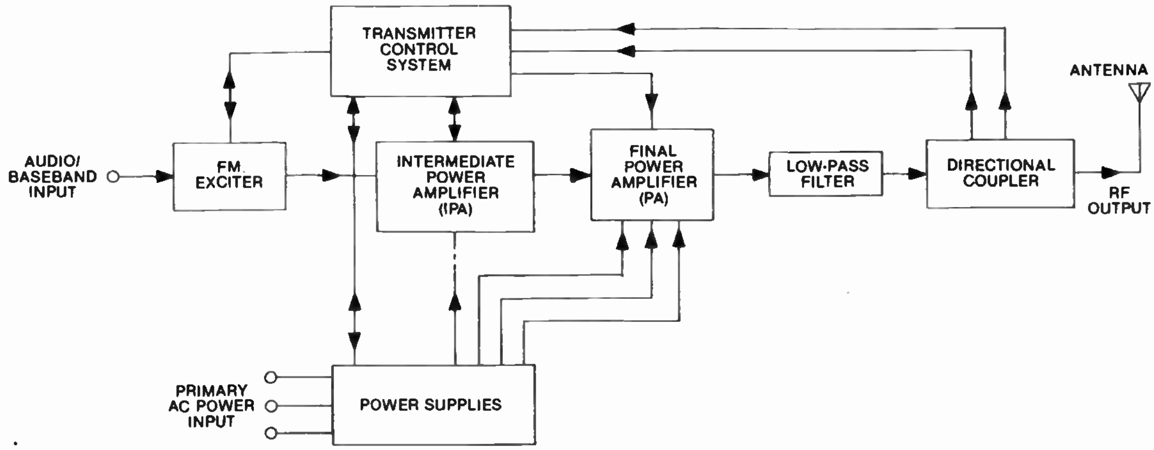


Figure 4.3-11. Simplified block diagram of an FM broadcast transmitter.

in the demodulated output at the receiver. A secondary effect of this distortion may include stereo crosstalk into the SCA, (subcarrier) signals.

The composite baseband signal is frequency modulated onto an RF carrier by the modulated oscillator (see Figure 4.3-12). Frequency modulation is achieved by applying the composite baseband signal to a voltage tunable RF oscillator. The modulated oscillator usually operates at the carrier frequency and is voltage tuned by varactor diodes operating in a parallel LC circuit.

To have perfect modulation linearity, the RF output frequency (F_c) must change in direct proportion to the composite modulating voltage (V_M) applied to the varactor diodes (C_v). This requirement implies that the capacitance of the varactor diodes must change as nearly the square of the modulating voltage as shown in following relationships:

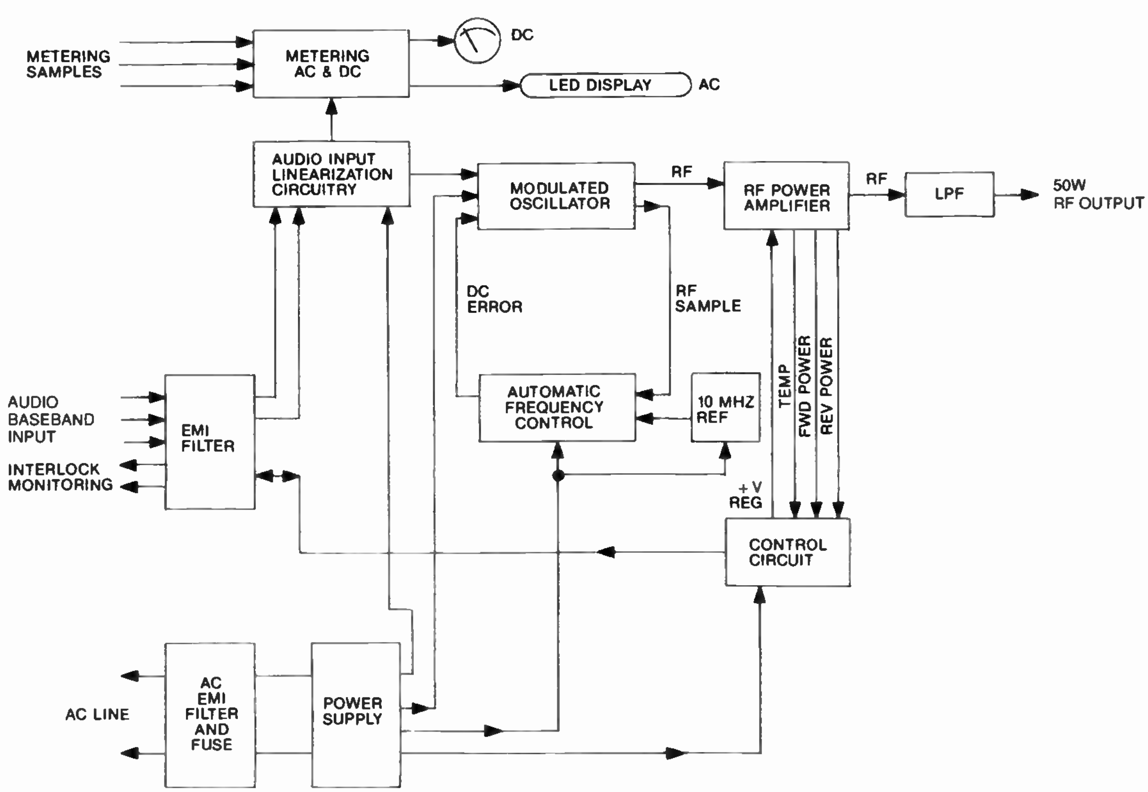


Figure 4.3-12. FM exciter block diagram using direct FM.

(F_c) is proportional to (V_M) if:

$$F_c = \frac{1}{2\pi\sqrt{(L)(C_t)}}$$

and:

$$C_v = \frac{K}{(V_m^2)} \qquad (\text{if } C_{\text{fixed}} = 0) \qquad (6)$$

(Desired linear voltage to frequency translation)

where:

F_c = Instantaneous carrier frequency
L = Inductance of resonant circuit
C_t = Total capacitance across L ($C_{\text{fixed}} + C_{\text{varactors}}$)
C_v = Capacitance of varactor tuning diodes
K = Varactor constant
V_M = Baseband modulating voltage

Unfortunately, the voltage versus capacitance characteristic of practical varactor diodes is not the desired square law relationship. All varactor-tuned oscillators have an inherently nonlinear modulating characteristic. This nonlinearity is very predictable and repeatable for a given circuit configuration, making correction by complementary predistortion of the modulating signal feasible. Suitable predistortion can be applied to the composite baseband signal by using a piece-wise linear approximation to produce the desired complementary transfer function. Figure 4.3-13 shows a typical network of switching diodes and resistors used for complementary predistortion of the composite baseband. Figure 4.3-14 shows how the predistortion network is cascaded with a nonlinear voltage-tuned oscillator to produce a linearized FM modulator.

It is also possible to improve both the linearity and S/N of the modulated oscillator by demodulating its RF output to baseband and then feeding some of this baseband with the proper phase relationship back into the composite input of the modulator. This type of configuration places the entire modulated oscillator within a negative feed back loop and transfers the responsibility for maintaining linearity to the demodulator. Digital demodulation schemes can be made very linear, but the additional complexity and the potential problems with loop stability have limited the applications of this approach to linearization.

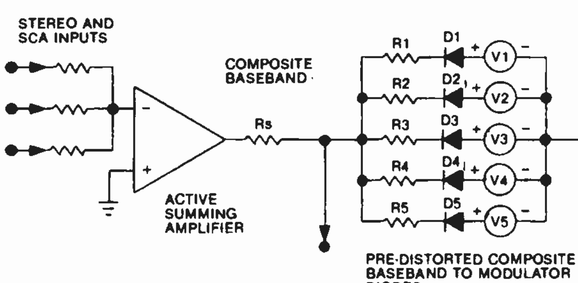


Figure 4.3-13. Predistortion network.

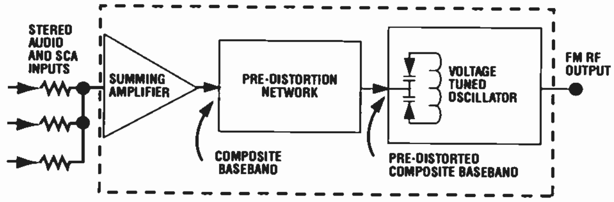


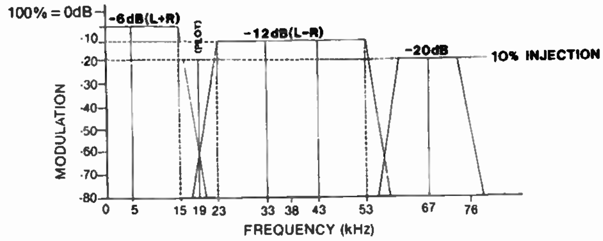
Figure 4.3-14. Linearized FM modulator block diagram.

Analog modulator linearization has reduced harmonic and intermodulation distortion to less than .01% in the current generation of equipment. Any distortion of the baseband signal caused by the modulated oscillator will have secondary effects on stereo and SCA crosstalk, which are quite noticeable at the receiver in spite of the rather small amounts of distortion to the baseband. For example, if the harmonic distortion to the baseband is increased from .05%–1.0%, as much as 26 dB of additional crosstalk into the SCA can be expected.

For illustrative purposes, Figures 4.3-15(a), 4.3-15(b) and 4.3-15(c) give representations of the fundamental and second order terms in the composite baseband spectrum with increasing amounts of harmonic distortion in the modulated oscillator. Figure 4.3-15(b) shows this spectrum after 0.05% harmonic distortion has been added to each component. Note that the second order stereo (L − R) sidebands are 78 dB below 100% modulation (about 58 dB below the 67 kHz SCA with a 10% injection). With normal (audio program) energy distribution in L − R and the SCA, crosstalk from stereo into the SCA will be more than 60 dB below the SCA subcarrier. Figure 4.3-15(c) shows the same baseband spectrum with 1.0% harmonic distortion. The second order stereo sidebands are now only 32 dB below the SCA, representing an increase of as much as 26 dB over the 0.05% distortion case, depending on the respective energy distributions in (L − R) and the SCA.

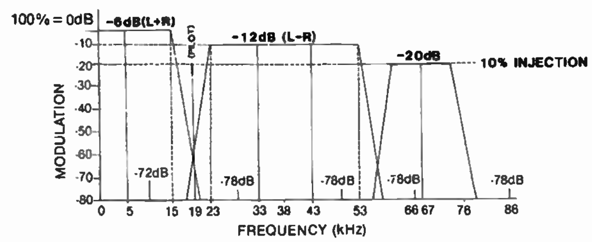
Transient intermodulation (TIM) distortion is usually not a factor in varactor-tuned modulated oscillators. The modulation bandwidth capability is generally more than ten times the composite bandwidth and no negative feed back is used to maintain linearity.

Ensuring that the composite baseband signal undergoes minimal distortion in the modulation process will suppress undesired harmonic and intermodulation products in the baseband, making the FM exciter transparent to the signals coupled into it. All exciter stages after the modulated oscillator operate as broad band amplifiers with minimal bandwidth limitations. The best analog FM exciter technology presently available is capable of transmitting near compact disk quality with less than .01% distortion and a S/N of better than 90 dB. Digital FM exciter technology utilizing DDS can now go beyond the limitations of analog modulators to provide the full 16bit digital performance of a compact disk.



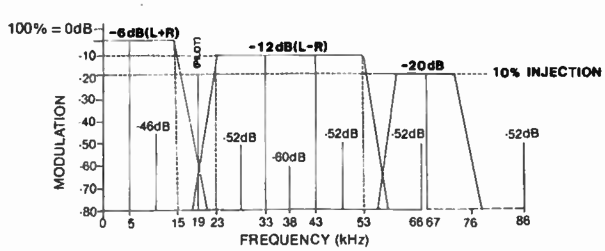
L OR R ONLY MODULATED 100% @ 5 kHz. UNMODULATED SCA @ 10% INJECTION

Figure 4.3-15(a). Ideal demodulated composite baseband spectrum with no modulator distortion.



L OR R ONLY MODULATED 100% @ 5 kHz. UNMODULATED SCA @ 10% INJECTION.
INTERFERING SECOND HARMONIC STEREO SIDEBANDS ARE 58 dB BELOW SCA.
ONLY FUNDAMENTAL AND SECOND HARMONIC TERMS ARE SHOWN.

Figure 4.3-15(b). Demodulated composite baseband spectrum with 0.05% harmonic distortion in modulator.



L OR R ONLY MODULATED 100% @ 5 kHz. UNMODULATED SCA @ 10% INJECTION.
INTERFERING SECOND HARMONIC STEREO SIDEBANDS ARE 32 dB BELOW SCA.
ONLY FUNDAMENTAL AND SECOND HARMONIC TERMS ARE SHOWN.

Figure 4.3-15(c). Demodulated composite baseband spectrum with 1.0% harmonic distortion in modulator.

Automatic Frequency Control

The frequency stability of direct FM oscillators is not good enough to meet the FCC frequency tolerance of ±2,000 Hz. This requires an automatic frequency control system (AFC) that uses a stable crystal oscillator as the reference frequency.

The modulated oscillator need not have good long-term stability since the AFC feed back loop will correct for long-term drift to keep the average carrier frequency within limits. The modulated oscillator does need excellent short-term (less than 1 sec) stability because the control loop time constant must be long enough so that the AFC circuit does not try to remove desired low frequency audio modulation. This means that the oscillator is essentially running open-loop at frequencies above a few hertz so that the noise performance of the modulator will also be determined by the short term stability characteristics of the oscillator.

Phase-Locked Loop Automatic Frequency Control

Phase-locked loop (PLL) technology has provided a means of precisely controlling the carrier's average frequency while permitting wide deviation of the carrier frequency at baseband modulating frequencies. This implies that a PLL system behaves like an audio high-pass filter with higher modulating frequencies being ignored by the control loop while lower frequencies are considered to be errors in the average frequency and are tracked out by the loop. An added advantage of the PLL is the ability to synthesize the desired frequency from a single reference oscillator, thereby eliminating the need to change crystals when changing the frequency of the exciter.

The block diagram shown in Figure 4.3-16 includes the key elements in the PLL. The output of the modulated oscillator operating at the carrier frequency is digitally divided down to a frequency of a few kilohertz or even less, called the comparison frequency. Likewise, the reference crystal oscillator is also digitally divided down to the reference frequency. The two frequencies are compared in a digital phase/frequency detector to develop an error voltage which corrects the carrier frequency of the modulated oscillator. The reason for dividing the modulated oscillator frequency so many times is to reduce the modulation index enough to limit the peak phase deviation at the reference frequency to a value that will not exceed the linear range of the phase/frequency detector. If the linear range is exceeded, the loop will lose lock. This is why some exciters may lose AFC lock in the presence of low frequency modulation components.

The phase detector output is integrated and low-pass filtered to remove the comparison frequency and all other frequency components above a few hertz so that the AFC circuit does not try to track-out low frequency modulation. Some FM exciters use a dual-speed PLL in order to keep the loop turn-over frequency low enough to maintain good amplitude and phase response at 30 Hz, while also providing quick lock-up time. The PLL error correction circuitry must respond quickly during the initial frequency scan of the FM band to achieve lock-up to the precision reference oscillator in a few seconds. The loop bandwidth is wide during acquisition and lock-up. After lock is achieved, the bandwidth is reduced to provide the optimum modulation characteristic.

The reference oscillator is usually temperature compensated and requires no warm-up to maintain ±3 PPM or better accuracy over the operating temperature range. Ten MHz is often selected as the reference frequency for convenient comparison to international znfrequency standards. For more information about PLL frequency synthesizers see [5].

Digital FM Exciter using Direct Digital Synthesis

DDS eliminates the need for a PLL in the FM modulation process. It does so, by directly synthesizing the carrier frequency, including FM modulation, from a

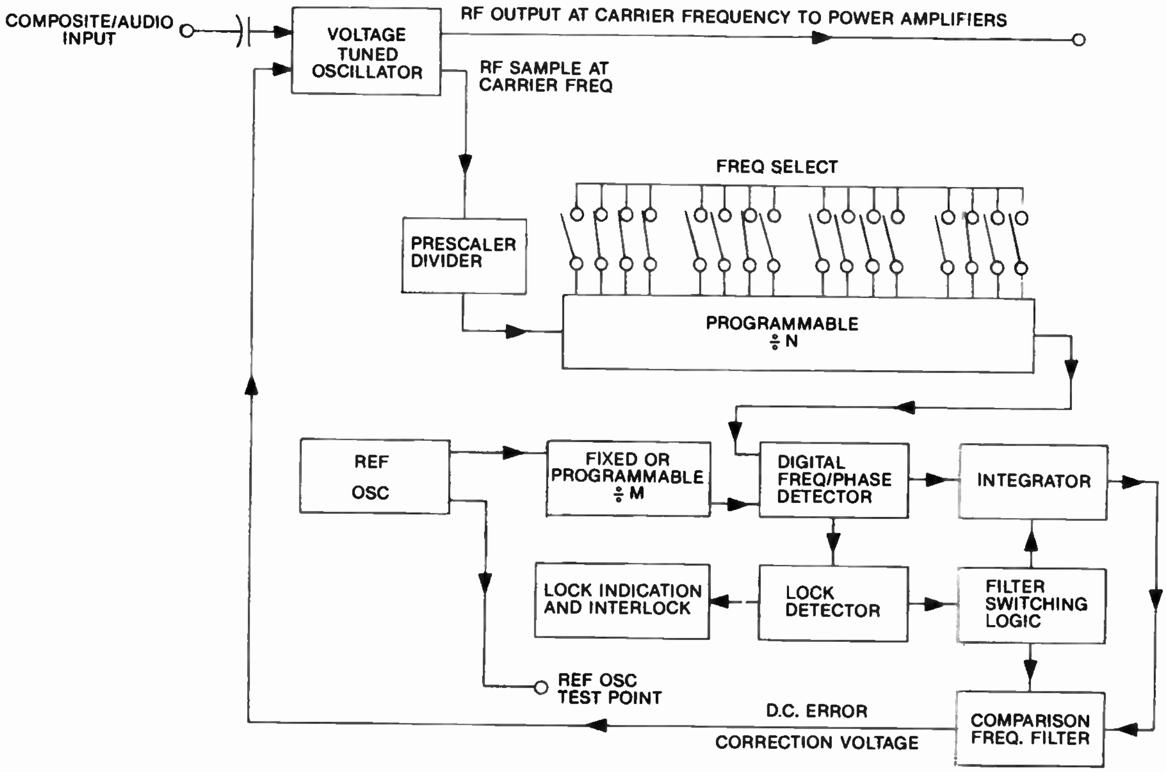


Figure 4.3-16. Phase-locked loop frequency synthesizer.

sinewave look-up table in a programmable read only memory device (PROM) operating in conjunction with a digital phase accumulator and a fast digital-to- analog (D/A) converter. When this technique is combined with DSP technology, the entire process of generating stereo baseband with SCAs and FM modulating this baseband information onto the RF carrier can be done entirely in the digital domain. The cost to performance ratio of DDS/DSP technology has made it competitive with the analog technology used in present exciters. The full benefit of DDS/DSP technology requires digital transmission of audio information as an uncompressed, digital, bit stream all the way from the digital audio source through a digital console, digital audio processing and an uncompressed, digital STL to the AES3 digital input port of the DSP/DDS exciter. This same technology is used in the fully digital audio broadcast (DAB) services including the Eureka 147 (EU-147) transmission standard and other technical standards presently being implemented worldwide.

With DDS, the completely modulated FM waveform, is generated totally in the digital domain. As digital modulation is an inherently linear process, no predistortion is required. The FM signal generated by a DDS device has extremely low noise and distortion for true 16 bit digital audio quality (−96 dB FM S/N and 0.0016% harmonic distortion for ±75 kHz deviation and 75 μsec preemphasis/deemphasis).

The current generation of DDS exciters use a 32 bit NCO. The basic resting frequency of the NCO is set by a 32 bit tuning word. Frequency modulation occurs when modulation data varies the structure of the tuning word within the phase accumulator section of the NCO. The modulated output of the NCO is converted to analog FM, upconverted, filtered and amplified to become the RF excitation for a conventional FM broadcast transmitter RF amplifier chain. A block diagram of a DDS digital FM exciter is shown in Figure 4.3-17.

DDS FM exciters also eliminate several basic limitations found in analog exciters using direct FM via the modulation of VCOs. Very low audio frequencies must be filtered from program signals to avoid affecting the AFC circuits of the analog exciter, which see very low modulating frequencies as an off-frequency condition that needs correction. A DDS based FM exciter has no such limitation and modulation response extends virtually to dc (zero hertz). These lower octaves of program material are important for sonic realism and to preserve the phase correlation existing in the original program.

Digital Modulator

The digital modulator which is the heart of a digital exciter, utilizes a 32 bit NCO to digitally generate the completely modulated FM waveform. Other supporting circuitry, including a digital peak detector, is used to drive the front panel modulation display. The block diagram in Figure 4.3-18 shows the functional subsystems of a digital modulator.

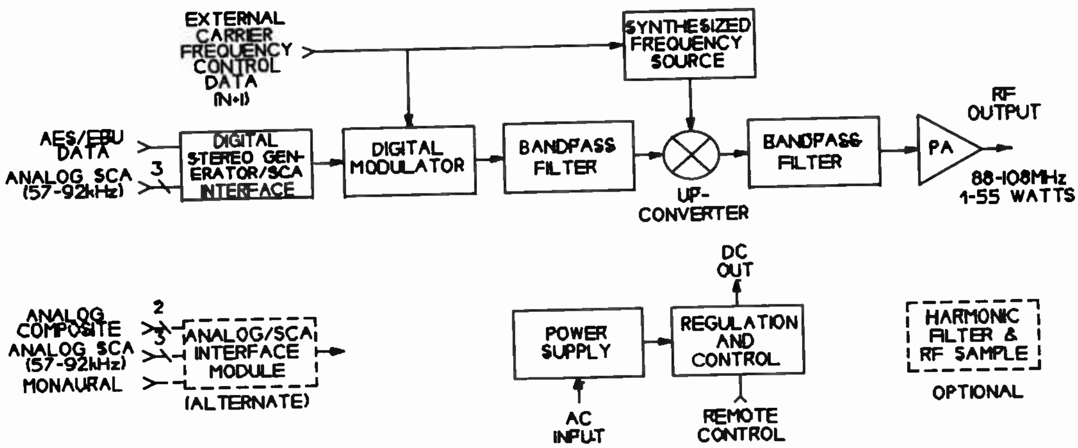


Figure 4.3-17. Harris DIGIT-CD DDS Digital FM exciter.

The input to the digital modulator is fed modulation data in the format and at the clock rate required. This data represents the stereo baseband created in the digital input module which contains the DSP digital stereo generator and subcarrier input circuitry.

The digital modulator module includes a precise, digital peak detector to provide the drive for the front panel peak modulation display. This circuit is driven by the same data as the NCO modulator. Therefore, the modulation indication on the front panel of the digital exciter has very high accuracy to within 0.25% of the true FM deviation value at any modulation index or frequency.

The output of the NCO is then D/A converted to a precise, conventional FM signal at an intermediate frequency and passed through a bandpass filter to remove the images produced in the DDS process.

Digital Input and DSP Stereo Generator

DDS techniques are compatible with standard AES3 digital audio, which provides the last link in maintaining a 100% digital audio path from the program source through the generation of the modulated FM carrier, with no intervening A/D or D/A conversions to add noise and distortion. The DDS exciter includes a built-in DSP stereo generator to convert the incoming AES3 digital stereo into the digital stereo modulation data needed for the NCO to generate FM stereo. Most DDS exciters also offer as an option a high-quality A/D converter that will convert the analog baseband output of an analog stereo generator into the digital data format needed by the NCO. The same high-quality modulation performance as for digital input signals is provided under these conditions. Auxiliary signals such as SCA and Radio Broadcast Data Service (RBDS)/RDS are accepted as modulated analog waveforms from external devices, then A/D converted to digital data and applied to the NCO for simultaneous transmission with normal stereo program material. Figure 4.3-19 shows the major signal blocks in the digital input module.

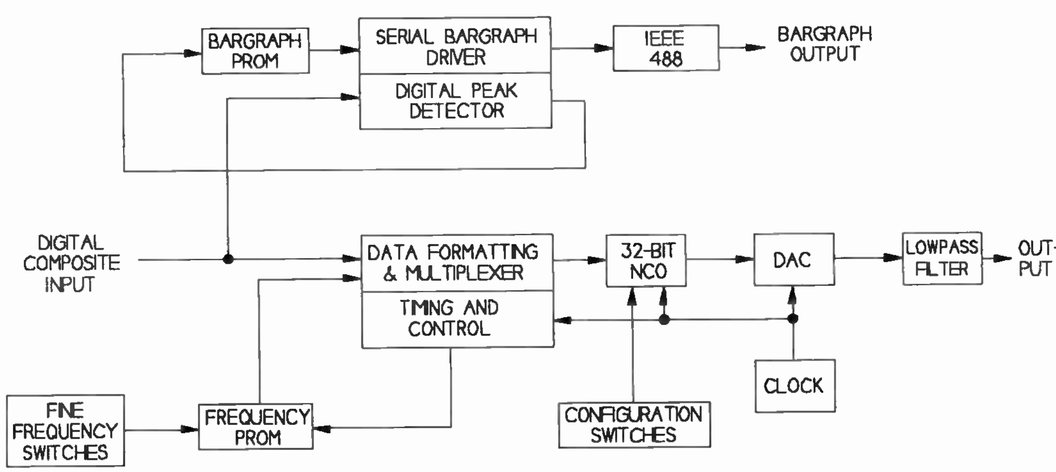


Figure 4.3-18. Digital Modulator in the Harris DIGIT-CD FM exciter.

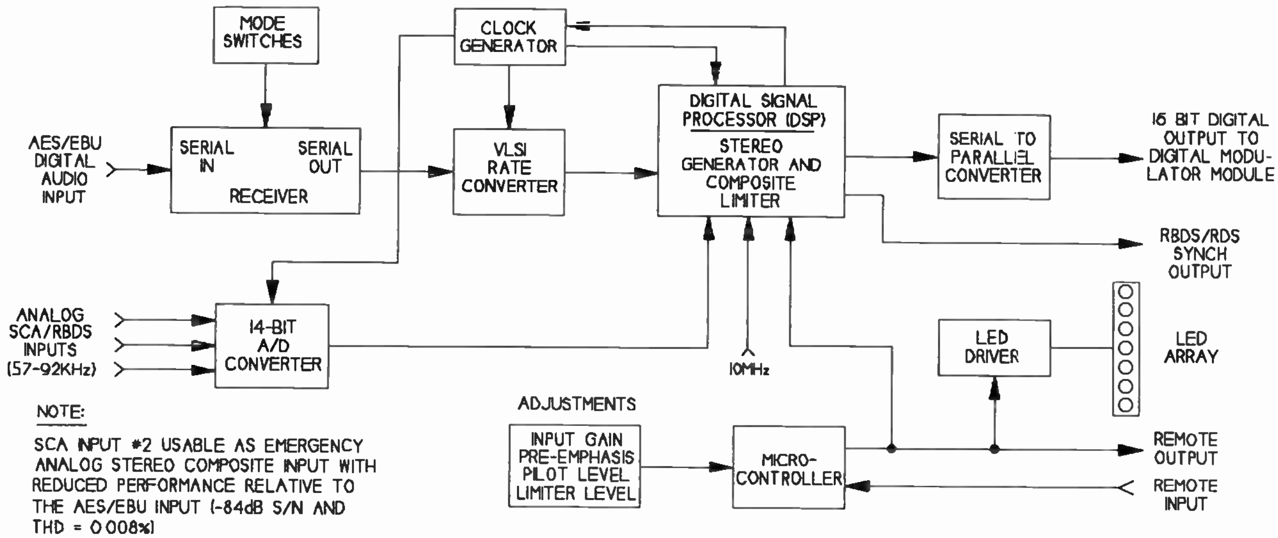


Figure 4.3-19. Harris DIGIT-CD Digital input module and DSP stereo generator.

The digital input module provides the interface, monitoring and control circuitry for these functions:

- Accepts standard AES/EBU stereo digital data at any rate from 20.8 kHz–56 kHz, normally 32 kHz. Rate conversion is automatic (self-clocking)
- Accepts up to three analog subcarriers in the range 57–92 kHz
- Converts all input signals to the composite digital format needed by the digital modulator
- Provides a 19 kHz output to synchronize an external RBDS/RDS generator
- Supplies digital composite limiting using a special implementation of "lookahead" digital technology
- Provides gain adjustment for summed SCA, overall deviation, limiting level and pilot level
- Provides switchable LED display showing either total peak deviation or limiting level
- Allows user selection of program channel mode (AES/EBU or analog), preemphasis on/off/time-constant, mono on/off and data error detection mode. For more detailed information about digital modulators and other aspects of digital FM exciter design and performance see [27, 30, 31, and 32].

FM by I/Q Modulation Techniques

Another method used to generate FM is the quasi digital, in-phase/quadrature (I/Q), RF modem technique. Two identical analog RF mixers are fed in-phase (I) and quadrature (Q) analog baseband signals derived from two separate D/A converters at the output of a DSP digital stereo generator. The (I) mixer is also fed the in-phase carrier frequency while the (Q) mixer is fed the carrier frequency phase shifted by 90° from the (I) mixer. When the RF outputs of these two mixers are summed the desired FM RF spectrum is produced if everything is perfectly balanced and matched. Practical limitations in the matching of these mixers require that feed back incorporating sophisticated nonlinear DSP adaptive correction be placed around these mixers and the (I) and (Q) (D/A) converters driving them. This technique has the advantage of producing FM at the carrier frequency instead of at an IF frequency but has limited modulation performance and spectral purity.

Exciter Metering

Metering of important operating parameters can be provided by a combination of analog metering and a digital LED display. Steady-state parameters are usually selected by a multiposition switch and displayed on an analog or LCD multimeter. Typical steady state functions include regulated, preregulated and unregulated supply voltages; the AFC control voltage; RF power amplifier collector voltage and current; forward output power and reflected power.

A color-coded peak reading display is usually provided to constantly monitor the peak FM deviation. A high-speed peak detector gives accurate peak readings on signals from dc to 100 kHz. A one-shot multivibrator circuit can provide a clear indication of short transient peaks exceeding 100% modulation. Digital exciters can directly read the peak values of the modulation data producing the FM deviation in the NCO.

Exciter Packaging

Protection of sensitive circuits within an FM exciter from external electromagnetic interference (EMI) is important because the unit is often located in the near field of multiple broadcast antennas operating over a broad range of frequencies. The exciter should be protected from conducted EMI by use of RC and/or LC filters on all leads entering the cabinet, including the ac line. Additionally, the power transformer may have an electrostatic shield between the primary and secondary windings. The modulated oscillator is often shock mounted to prevent the transmission of mechanical vibrations from the transmitter's blower. This avoids microphonic pick-up by the modulator that would degrade the FM S/N. Magnetic shielding of the analog FM modulator is also used to prevent hum pick-up from nearby transformers. In some cases, a hum-bucking circuit is provided to help cancel hum induced into the analog modulator.

The mechanical construction of most present day exciters is designed around a plug-in modular or semi-modular approach, which allows easy removal of subassemblies for repair or replacement.

The exciter chassis may be mounted on pull-out slides so that all subassemblies are accessible while the unit continues to operate.

Exciter Output Stage

The broad band RF amplifier in the exciter amplifies the output of the modulated oscillator from a power level of a few milliwatts up to an output level in the range of 5–50 W. The output stage is usually protected against damage that could be caused by an infinite voltage standing wave ratio (VSWR) (on the output) at any phase angle.

The typical RF amplifier is designed to have a bandwidth of at least 20 MHz, using successive broad band impedance matching sections for each stage. Each group of matching sections consists of micro strip or lumped elements.

The broad band performance of the RF amplifier eliminates the need for adjustments to any particular frequency within the FM band. The exciter output is transparent to the signal generated by the modulated oscillator, and the amplifier stability is enhanced under varying load conditions.

A micro strip directional coupler is often incorporated in the RF amplifier output network. This coupler supplies information to the exciter control circuitry which provides automatic control of power output level and provides protection against operation under high VSWR conditions.

All current-generation FM exciters typically produce at least 50 W of RF output so they can be used as a complete transmitter for educational stations (with the addition of a harmonic filter at the output). For higher power level requirements, the exciter is used to drive an external power amplifier.

RF Power Amplifiers

The remainder of the FM transmitter consists of a chain of power amplifiers, each having from 8–20 dB of power gain. Ideally, the transmitter should have as wide a bandwidth as practical with a minimum of tuned stages. Broad band solid-state amplifiers are preferred to eliminate tuned networks in the RF path. Higher powered transmitters in the multikilowatt range may use multiple tube stages each with fairly low gain such as in the grounded grid configuration or a single grid driven power amplifier (PA) stage with high gain and efficiency. The cost, redundancy and wide bandwidth benefits of solid-state transmitters has made them attractive at power levels up to 20 kW. At higher power levels, beyond 20 kW, the lower cost per watt of high power, single-tube transmitters still make them attractive even though the modulation performance and reliability is less than that of a solid-state transmitter. Design improvements in tube-type power amplifiers have concentrated on improving bandwidth, reliability and cost effectiveness while design improvements in solid-state amplifiers have focused on continuous cost reduction to make them competitive with tube technology at ever increasing power levels.

RF Power Amplifier Performance Requirements

The basic function of the power amplifier is to amplify the power of the exciter output to the desired transmitter power output level. Most of the overall transmitter performance characteristics are determined by the exciter, but a few are established or affected by the following power amplifier characteristics:

- The RF output level at harmonics of the carrier frequency is almost completely a function of the attenuation provided by the power amplifier output matching circuit and output low-pass/notch filters. The FCC limit in decibels is [43 dB + 10 · log (pwr in watts) dB] or 80 dB whichever is less (73 dB for 1 kW output or 80 dB for 5 kW and higher)
- The major source of AM noise usually originates in the last power amplifier stage. The FCC limit is 50 dB below 100% equivalent AM modulation.
- The RF power output control system which must keep the output within +5% and −10% of authorized output is usually achieved in the final power amplifier
- Inadequate power amplifier RF bandwidth, particularly with respect to phase linearity (constant time delay) across the signal bandwidth, can reduce stereo separation and cause SCA crosstalk
- The presence of standing waves on the transmission line from the power amplifier to the antenna may also interact with the power amplifier to cause degraded stereo separation and SCA crosstalk.

The power amplifier should provide trouble-free service and be easy to maintain and repair. Good overall efficiency is also desirable to reduce the primary power consumption and heat load released into the transmitter room.

Power Amplifier Bandwidth Considerations

As mentioned earlier, the FM signal theoretically occupies infinite bandwidth. In practice, however, truncation of the insignificant sidebands (typically less than 1% of the carrier power) makes the system practical by accepting a certain degree of signal degradation. The transmitter power amplifier bandwidth affects the modulation performance. Available bandwidth determines the amplitude response and group delay response. There is a trade-off involved between the bandwidth, gain and efficiency in the design of a power amplifier.

The bandwidth of an amplifier is determined by the load resistance across the tuned circuit and the output or input capacitance of the amplifier. For a single-tuned circuit, the bandwidth is proportional to the ratio of capacitive reactance to resistance:

$$BW \propto \frac{K}{2\pi f_c R_L(C)} = \frac{K(X_C)}{R_L} \tag{7}$$

where:

BW = Bandwidth between half-power points (BW_{3dB})
K = Proportionality constant
R_L = Load resistance (appearing across tuned circuit)
C = Total capacitance of tuned circuit (includes stray capacitances plus output or input capacitances of the tube)
X_C = Capacitive reactance of C
f_c = Carrier frequency

The load resistance is directly related to the RF voltage swing on the tube element. For the same power and efficiency, the bandwidth can be increased if the capacitance is reduced.

Effects Of Circuit Topology And Tuning On FM Modulation Performance

FM broadcast transmitter RF power amplifiers are typically adjusted for minimum synchronous AM (incidental amplitude modulation) which results in symmetrical amplitude response. This ensures that the transmitter's amplitude passband is properly centered on the FM channel. The upper and lower sidebands will be attenuated equally or symmetrically which is assumed to result in optimum FM modulation performance. This is true only if the RF power amplifier circuit topology results in simultaneous symmetry of both amplitude and group delay responses.

The tuning points for symmetrical amplitude response and symmetrical group delay response usually do not coincide, depending on the circuit topology. Therefore, simply tuning for minimum synchronous AM (symmetrical amplitude response) does not necessarily result in best FM modulation performance.

Actually, symmetry of the group delay response has a much greater effect on FM modulation distortion than does the amplitude response. Tuning for symmetrical group delay will cause the phase/time delay errors to affect the upper and lower sidebands equally or symmetrically. The group delay response is constant if the phase shift versus frequency is linear. In this case, all components of the signal are delayed in time, but no phase distortion occurs.

Measurements taken on a typical FM transmitter as well as computer simulations showed that tuning the RF power amplifier for symmetrical group delay response resulted in minimum distortion and crosstalk. [23] It confirmed that group delay response asymmetry causes higher FM modulation distortion and crosstalk than amplitude response asymmetry. The transmitter should be tuned for symmetrical group delay response which results in best FM modulation performance rather than symmetrical amplitude response which results in minimum synchronous AM.

Power Amplifier Output Source Impedance

At the milliwatt levels used in RF test equipment, it is customary to use 50 Ω as the standard impedance. Consequently, test equipment is typically designed to provide both 50 Ω source and load impedances to systems under test. This approach minimizes any reflections to systems under test since both the transmitter (source) and the termination (load) will absorb reflected energy. A 50 Ω source impedance is usually provided by placing a 50 Ω build out resistor in series with a low impedance voltage source (Thévenin equivalent). The closed circuit voltage with this configuration is exactly ½ of the open circuit voltage, meaning that half of the total available RF power is dissipated in the source resistance. The best possible efficiency for this system is 50% assuming the voltage source is 100% efficient without the source resistance.

It becomes obvious that while an FM transmitter is designed to drive a 50 Ω load, it does not itself have an output source impedance of 50 Ω. In order to achieve high efficiency, the transmitter must have a very low output source impedance so that nearly all of the power is delivered to the load. The plate dissipation indirectly represents some of the power lost within the low source resistance. Since the low source impedance of the transmitter provides a mismatch to reflected power from the load, this power is almost totally reflected back from the transmitter output stage toward the load again.

Intermediate Power Amplifiers

The IPA is located between the exciter and the final amplifier in higher power transmitters that require more than about 50 W of drive to the final amplifier. The IPA may consist of one or more tubes or solid-state amplifier modules.

Interstage Coupling Circuits

The separate IPA output circuit and the final amplifier input circuit are often coupled together by a coaxial transmission line. Impedance matching is usually accomplished at either end by one of the configurations shown in Figures 4.3-20(a), 4.3-20(b), 4.3-20(c) and 4.3-20(d).

The interconnecting transmission line between the coupling circuits should be properly matched to avoid a high VSWR. Directional wattmeters are normally placed in the line to measure forward and reflected power from which standing wave ratio can be established. The VSWR is established by the match at the load end of the transmission line.

Solid-state RF power devices possess a very low load impedance at the device output terminal, so that an impedance transformation that goes through the 50 Ω intermediate impedance level is required to couple these devices into the relatively high impedance of the final amplifier grid circuit. Therefore, virtually all solid-state IPA systems have a 50 Ω impedance point within the system that can be used to feed the antenna in an emergency.

The tube in the final amplifier stage of most newer design high power transmitters only requires between 150–600 W of drive. This permits the use of solid-

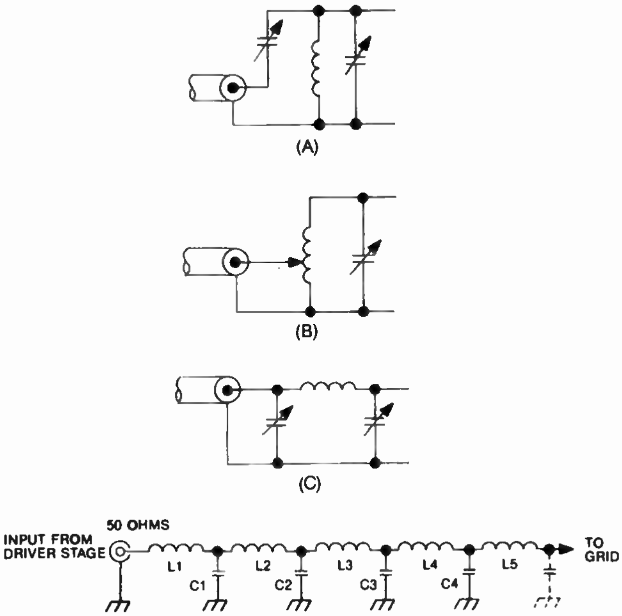


Figure 4.3-20(a–d). Interstate RF coupling circuits.

state, wideband, power amplifier modules to boost the exciter's power up to the level required to drive the grid of the final tube.

Solid-State RF Power Amplifier Systems

A solid-state RF power amplifier almost always consists of a system of individual amplifier modules that are combined to provide the desired power output. The advantages of using several lower power modules instead of a single high power amplifier include:

- Redundancy is provided by isolating the input and output of each module, permitting uninterrupted operation at reduced power if one or more of the modules fails.
- The ability to repair or replace failed modules without having to go "off-the-air"
- More effective cooling of each power device by splitting the concentration of heat to be dissipated into several areas instead of one small area
- Better isolation between the amplifier modules and the input circuit of the final power amplifier or antenna is provided by the combiner/isolator
- Redundant power supplies and air cooling systems for each module improve overall reliability.

Each RF power amplifier module consists of one or more solid-state devices with broad band impedance transformation networks for input and output matching. A new generation of Metal Oxide Semiconductor Field Effect Transistor (MOSFET) devices permits the design of broad band amplifier stages that exhibit both high efficiency and the wide bandwidth necessary to cover the FM broadcast band.

The input impedance to the solid-state device is always lower than the desired 50 Ω input impedance, so a broad band impedance transformation scheme is required. This is usually accomplished by a combination of coaxial baluns and push-pull coaxial line sections that are cross-coupled to provide 4:1 or higher transformation ratios over the FM band.

By operating two devices in push-pull, the input impedance (differential) is double that of a single ended circuit, and the suppression of even order harmonics is enhanced. Two devices fed in this manner also provide some degree of redundancy within the module itself since partial RF output can be obtained with one device failed. In a similar manner, the low output impedance of these solid-state devices can be transformed up to the desired 50 Ω module output impedance where combining occurs. Figure 4.3-21(a) illustrates a simplified schematic of a broad band,

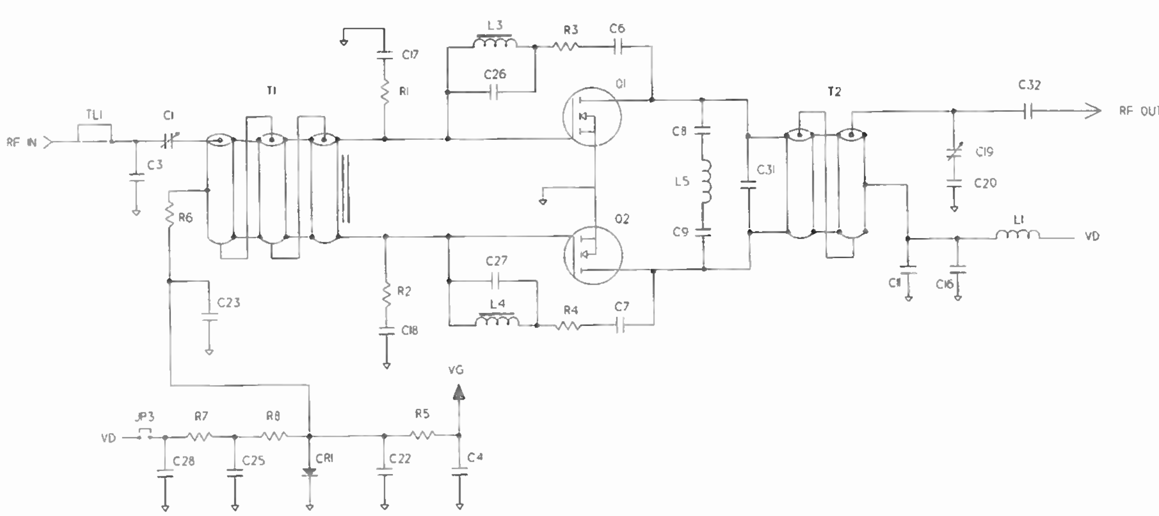


Figure 4.3-21(a). Schematic of broadband, 425 W, MOSFET, RF power amplifier module.

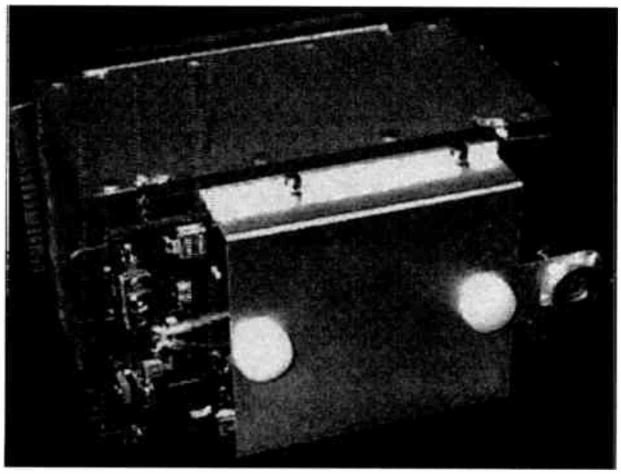


Figure 4.3-21(b). Broadband, 850 W, MOSFET, RF power amplifier module.

425 W, MOSFET, RF amplifier module utilizing the push-pull configuration. Figure 4.3-21(b) is a photograph of this RF amplifier module.

Solid-State Amplifier Splitting and Combining

Two frequently used types of splitting/combining schemes are:

- A 90° hybrid splitter or combiner ("N-1" hybrids required to split or combine "N" inputs). (See the section on transmitter output combining.)
- A Wilkinson "N-way" in-phase splitter or combiner.

Either type of splitter/combiner must provide isolation between the individual power amplifier modules and low loss splitting or combining of the total power.

The cascaded 90° hybrid system shown in Figure 4.3-22 provides double isolation between the power amplifiers and the load by first combining the two pairs of amplifiers and then combining the outputs of the first two combiners. A portion of the reflected power, caused by a mismatch at the output, will be dissipated in the reject loads so that the power amplifier modules will operate into a lower VSWR than exists at the output. The unbalanced 50 Ω reject loads are accessible for monitoring of reject load power which is useful in determining the balance of the system.

Wilkinson system shown in Figure 4.3-23 is a simple and effective way to split and combine modules operating in phase, but usually requires a balanced reject load making reject power measurements more difficult. By adding additional coaxial balun sections to the Wilkinson (Wilkinson/Gysel), it is possible to use unbalanced reject loads.

Adaptive Control of the Combiner Configuration

Both the 90° hybrid and Wilkinson combining systems require resistive RF power reject loads to provide isolation between the amplifier modules in the event that one or more of the modules fail. A portion of the RF power from the remaining modules is wasted in the reject loads instead of being delivered to the output. Recent developments have made it possible to have a microcomputer monitor the degree of imbalance in the system and adaptively change the configuration of the combiner to losslessly compensate for the failure of one or more power amplifier modules. This is accomplished by having the microcomputer substitute the appropriate reactances in place of the resistive reject loads to maintain enough isolation for the remaining power amplifiers to work efficiently. This technique is used in the Harris "Z" plane combiner.

Since most splitter/combiner systems are designed around a 50 Ω input and output impedance level, these systems can be easily used as a low power stand-by transmitter by routing the output to the antenna system.

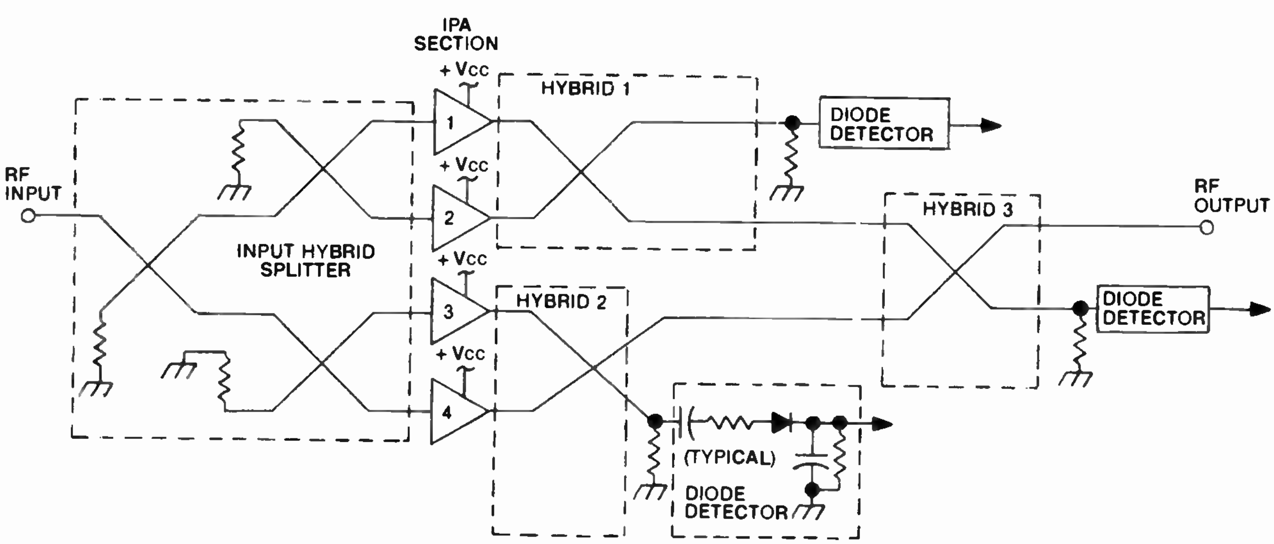


Figure 4.3-22. Cascaded 90° hybrid splitting/combining system.

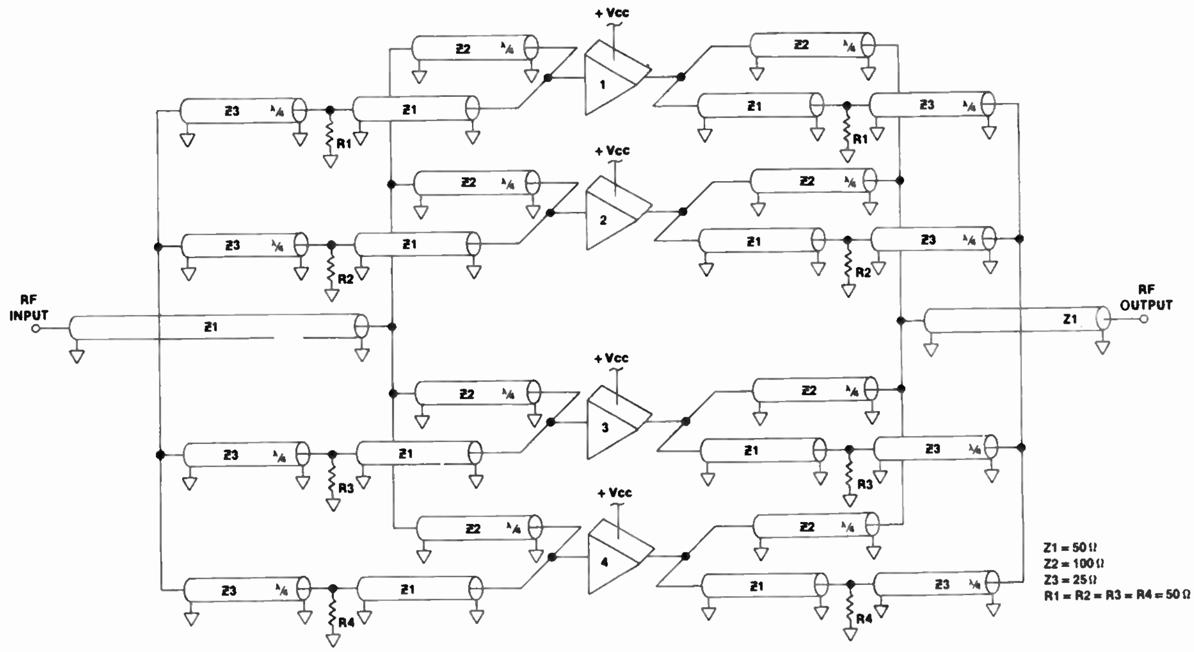


Figure 4.3-23. Wilkinson/Gysel in-phase splitting/combining system with unbalanced reject loads.

An RF low pass filter (LPF) is required only when directly feeding the antenna system. The harmonic suppression of the IPA is not as critical when driving a nonlinear power amplifier that also generates harmonics, because this stage will have its own LPF.

SOLID-STATE FM BROADCAST TRANSMITTERS

The techniques used to construct IPA systems can also be used to construct a completely solid-state transmitter using arrays of combined modules for the final output stage. An additional RF low pass filter is usually required to meet FCC emission requirements.

Advantages of Solid-State Transmitters

The primary advantages of a solid-state transmitter are the built-in amplifier and power supply redundancy, superior FM modulation performance, the ability to cover the entire FM band without the need for re-tuning, and elimination of tube replacement costs. A tubeless transmitter is nearly maintenance free.

Solid-State Transmitter Design Considerations

Several manufacturers offer solid-state FM broadcast transmitters with power outputs ranging from 100 W up to 20 kW, but present economic factors still favor the single tube FM transmitter for power levels above 20 kW.

In order for a solid-state transmitter to be cost and power consumption competitive with a single tube transmitter, the efficiency of the solid-state RF power amplifiers and combining system have to approach the 80% efficiency obtainable from tube type RF amplifiers. This high efficiency has been achieved with MOSFET solid-state devices at VHF frequencies.

Solid-state designs can provide high ac/RF efficiencies, with up to 80% dc to RF efficiency at the MOSFET device level and over 62% overall efficiency, from ac line in to RF output. This is actually better ac to RF efficiency at the 5 kW level than a typical single tube transmitter, which is normally less than 60%.

Some solid-state designs have added a few percentages to their overall ac to RF efficiency by optimizing their RF circuits over narrow band sections of the FM band. This approach is beneficial to the user who is certain there will not be a need to change the transmitter's frequency, or who is prepared to provide the transmitter modifications needed to do so.

Trends in the newest solid-state FM transmitters are to supply redundant RF, power supply and control circuits so as to keep the transmitter on the air at reduced power in the event one or more components should fail. Identical and interchangeable IPA and PA modules offer additional redundancy. RF modules that can be removed and inserted in an operating transmitter also provide the advantage of not requiring an off-air period for some maintenance services.

Solid-state transmitter layouts using direct, cable-free connection of the RF modules to the RF combiner have also been introduced and further enhance transmitter reliability and stability. Another enhancement provided in some current solid-state FM transmitters is an advanced, microprocessor based control system that can monitor detailed parameters within the transmitter and provide "intelligent control" of the transmitter system, including the RF combiner. This maximizes output power and minimizes reject load power under

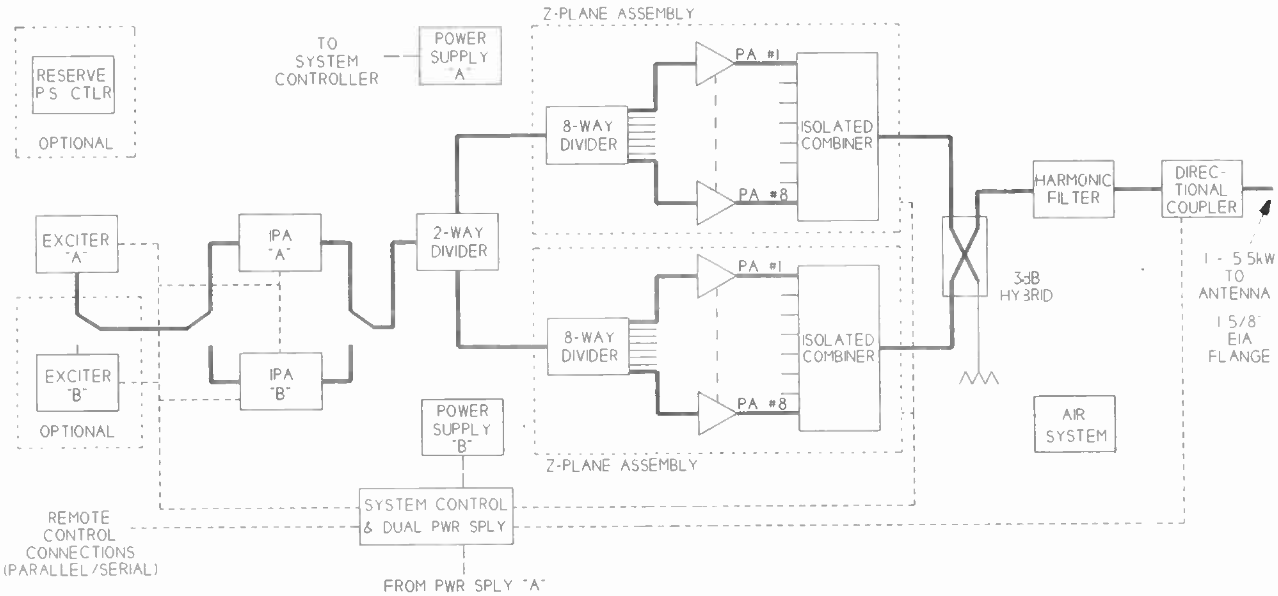


Figure 4.3-24. Harris "Z" 5 kW solid-state FM transmitter.

various combinations of active and inactive modules. Figure 4.3-24 shows a block diagram of a 5 kW solid state transmitter.

Solid-State Versus Single Tube Transmitter Purchase Cost

There is no engineering reason why a solid-state FM transmitter of 20–40 kW output power cannot be produced. The only economic factor at the present time is the cost premium, which tends to rise linearly with power for solid-state transmitters, while leveling off exponentially for high-power single tube transmitters. Figure 4.3-25 graphically shows that at power levels below approximately 8 kW, solid-state transmitters are less expensive to purchase than transmitters that use a tube. As the power level increases, the cost of a single tube transmitter increases incrementally, but not in direct proportion to the power increase. If the power output of a solid-state transmitter is doubled, it requires twice as many power amplifiers, power supplies, combiner ports etc., so that the cost very nearly doubles.

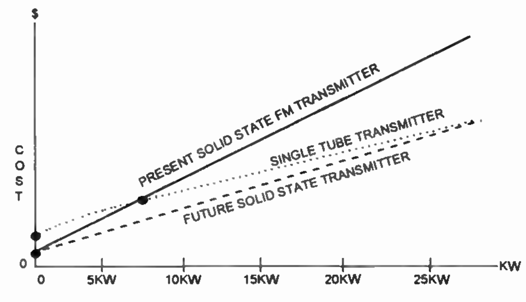


Figure 4.3-25. Solid-state versus single tube transmitter purchase cost.

Since the cost of a solid-state transmitter increases proportionately to the power output, the only way to make solid-state technology cost competitive with high power tube technology is to reduce the slope of the solid-state curve by driving the cost of solid-state devices down while simultaneously increasing the power density per RF device. This trend will continue, eventually making high power solid-state transmitters more economical to purchase.

Purchase price alone does not determine the overall economics of buying a solid-state FM transmitter. When the reduced cost of installation, maintenance, elimination of PA tube replacement cost and the overall cost of ownership are included in the comparison, a high-power, solid-state FM transmitter can become a good value. [6, 7, 8, and 9]

VACUUM TUBE POWER AMPLIFIER CIRCUITS

The amplitude of an FM signal remains constant with modulation so that efficient, nonlinear, Class C, amplifiers can be used.

FM broadcast vacuum tube power amplifier circuits have evolved into two basic types. One type uses a tetrode or pentode tube in a grid driven circuit while the other uses a high-mu triode in a cathode driven (grounded grid) circuit.

Cathode Driven Triode Amplifiers

The high-mu triodes used in cathode driven (grounded-grid) FM amplifiers were originally developed for linear single side band (SSB) amplifiers. Their characteristics are well adapted to FM broadcast use because the circuit is very simple, and no screen or grid bias power supplies are required. Figure 4.3-26 shows the basic circuit configuration. In this case, the

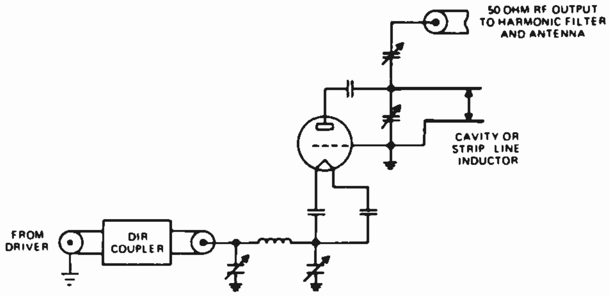


Figure 4.3-26. Cathode driven triode power amplifier.

grid is connected directly to chassis ground. Direct current grid current is the difference between dc cathode current and dc plate current. The output tank circuit is a shorted coaxial cavity which is capacitively loaded by the tube output and stray circuit capacitance. A small capacitor is used for trimming the tuning, and another small variable capacitor is used for adjusting the loading. A pi-network matches the 50 Ω input to the tube cathode.

The triodes are usually operated in the less efficient, Class B mode in order to achieve maximum power gain, which is on the order of 20 (13 dB). They can be driven into high efficiency, Class C operation by providing negative grid bias. This increases the plate efficiency but also requires increased drive power.

Most of the drive power into a grounded-grid amplifier is fed through the tube and appears in the stage's output. This increases the apparent efficiency so that the efficiency factor given by the manufacturer may be higher than the actual plate efficiency of the tube. The true plate efficiency is determined by dividing the output power by the total input power, which includes both the dc plate input power (Ip × Ep) and the RF drive power. Since most of the drive power is fed through the tube, any changes in loading of the output circuit will also affect the input tuning and driver stage.

With RF drive voltage on the cathode (filament) of the tube, some means of de-coupling must be used to block it from the filament transformer. One method employs high current RF chokes since the inductance can be very low at this frequency range. The other commonly used method feeds the filament power through the input tank circuit inductor.

Cathode driven stages are normally used only for the higher power stages. The first stage in a multitube transmitter is nearly always a tetrode because of its higher power gain.

Grounded-Grid Versus Grid Driven Tetrode Operation

Tetrodes may also be operated in the grounded-grid configuration by placing both the control grid and the screen grid at RF ground. Higher efficiency and gain can be achieved by placing negative bias on the control grid while placing a positive voltage on the screen grid of a cathode driven tetrode.

The input capacitance of a tetrode in a grounded-grid configuration is much less than a grid driven configuration and the input impedance is lower providing better bandwidth.

Approximate input capacitances of typical tubes:

	C input (pf)	
Tube Type	Grounded Grid	Grounded Cathode
4CX 3,000A	67	140
4CX 3,500A	59	111
4CX 5,000A	53	115
4CX15,000A	67	161
4CX20,000A/8990	83	190

Typical drive power requirements, as a function of plate voltage, for a 5 kW power amplifier:

	4500 V	5200 V
Configuration	Drive Power	Drive Power
Grounded Grid	340 W	280 W
Grid Driven	190 W	140 W

There are several trade-offs between the performance of grounded-grid and grid driven operations of a tetrode PA with respect to gain, efficiency, amplitude bandwidth, phase bandwidth and synchronous AM under equivalent operating conditions:

- When driving the PA into saturation, the bandwidth of the PA is limited by the output cavity bandwidth in the grounded-grid amplifier. The PA bandwidth in the grid driven amplifier is limited by the input circuit "Q," which is basically determined by the amount of swamping resistance
- Output bandwidth under saturation can be improved in either configuration by reducing the plate voltage. This involves a trade-off in efficiency with a smaller voltage swing. The bandwidth improvement can be obtained with a loss of PA gain and efficiency
- A grounded-grid saturated PA improves bandwidth over a grid driven saturated PA at the expense of amplifier gain. Best performance for FM operation is obtained when the amplifier is driven into saturation where little change in output power occurs with increasing drive power. Maximum efficiency also occurs at this point
- The phase linearity in the 0.5 dB bandwidth is better in a grid driven configuration. The Class C, grounded-grid, PA exhibits a more nonlinear phase slope within the passband, yet has a wider amplitude bandwidth. This phenomenon is due to interaction of the input and output circuits because they are effectively connected in series in the grounded-grid configuration. The neutralized, grid driven, PA provides more isolation between these networks, so they behave more like independent filters.

Grid Driven Tetrode and Pentode Amplifiers

A small tetrode tube such as the 4CX250B is commonly used as the only amplifier stage in 250 W transmitters and as the driver for higher power stages.

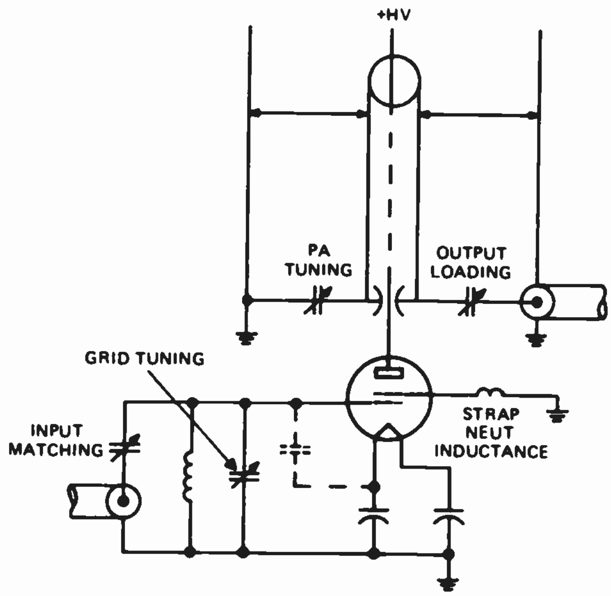


Figure 4.3-27. Grid driven grounded screen tetrode power amplifier.

Transmitters using tetrode amplifiers throughout usually have one less stage than those using triodes. Since tetrodes have higher power gain, they are driven into Class C operation for high plate efficiency. Against these advantages is the requirement for neutralization, along with screen and bias power supplies.

Figure 4.3-27 shows a schematic of a grid driven tetrode amplifier. In this example, the screen is operated at dc ground potential and the cathode (filament) is operated below ground by the amount of screen voltage required. This grounded-screen operation has the advantage of eliminating stability problems due to undesired resonances in the screen bypass capacitors. With directly heated tubes, it is necessary to use filament bypass capacitors. During grounded-screen operation, these bypass capacitors will need to have a higher breakdown voltage rating since they will have the dc screen voltage across them. The filament transformer must have additional insulation to withstand the dc screen voltage. The screen power supply provides a negative voltage in series with the cathode to ground and must have the additional capacity to handle the sum of the plate and screen currents. A coaxial cavity is used in the output circuit so that the circulating current is spread over large surfaces, keeping the losses very low. This cavity is a shorted quarter-wavelength transmission line section which resonates the tube's output capacitance. The length is preset to the desired carrier frequency, and then a small value variable capacitor is used to trim the system to resonance. Capacitive output coupling is used from the high RF voltage point to the 50 Ω transmission line. The 50 Ω input is capacitively coupled into the grid circuit inductor to provide the correct impedance match.

Pentode amplifiers have even higher gain than their tetrode counterparts. The circuit configuration and bias supply requirements for the pentode are similar to the tetrode since the third (suppressor) grid is tied directly to ground. The additional isolating effect of the (suppressor) grid eliminates the need for neutralization in the pentode amplifier.

Impedance Matching Into The Grid

The grid circuit is usually loaded (swamped) with added resistance. The purpose of this resistance is to broaden the bandwidth of the circuit by lowering the circuit "Q" and to provide a more constant load to the driver. It also makes neutralizing less critical so that the amplifier is less likely to become unstable with varying output circuit loading.

Cathode or filament lead inductance from inside the tube, through the socket and filament capacitors to ground can heavily load the input circuit. This is caused by RF current flowing from grid to filament through the tube capacitance and then through the filament lead inductance to ground. An RF voltage is developed on the filament which in effect causes the tube to be partly cathode driven. This undesirable extra drive power requirement can be minimized by series resonating the cathode return path with the filament bypass capacitors or by minimizing the cathode to ground inductance by using a specially designed tube socket using thin-film dielectric "sandwich" capacitors for coupling and bypassing.

High-power, grid driven, Class C, amplifiers require a swing of several hundred RF volts on the grid. To develop this high voltage swing, the input impedance of the grid must be increased by the grid input matching circuit. Since the capacitance between the grid and the other tube elements may be 100 picofarads (pF) or more, the capacitive reactance at 100 MHz will be very low unless the input capacitance is parallel resonated with an inductor. Figure 4.3-28(a) and 28(b) show two popular methods of resonating and matching into the grid of a high power tube. Both methods can be analyzed by recognizing that the desired impedance transformation is produced by an equivalent "L" network.

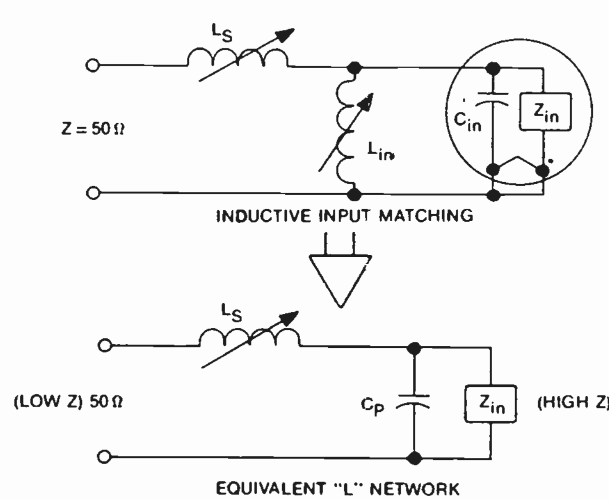


Figure 4.3-28(a). Inductive input matching.

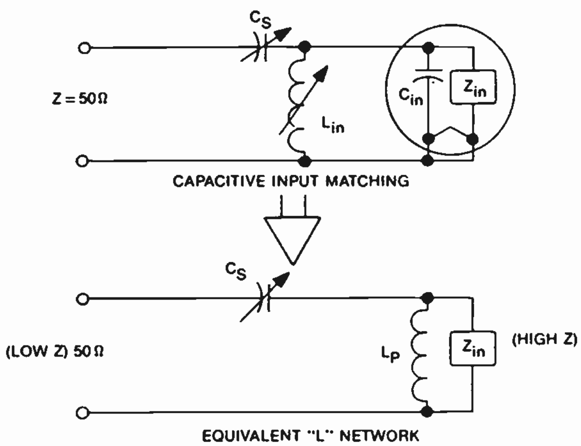


Figure 4.3-28(b). Capacitive input matching.

In Figure 4.3-28(a), a variable inductor (L_{in}) is used to raise the input reactance of the tube by bringing the tube input capacitance (C_{in}) almost to parallel resonance. Parallel resonance is not reached since a small amount of parallel capacitance (C_p) is required by the equivalent "L" network to transform the high impedance (Z_{in}) of the tube down to a lower value through the series matching inductor (L_s). This configuration provides a low-pass filter by using part of the tube's input capacitance to form (C_p).

Figure 4.3-28(b) uses variable inductor (L_{in}) to take the input capacitance (C_{in}) past parallel resonance so that the tube's input impedance becomes slightly inductive. The variable series matching capacitor (C_s) forms the rest of the equivalent "L" network. This configuration is a high-pass filter.

Neutralization

Cathode driven, grounded-grid amplifiers utilizing triodes do not require neutralization. The grid-to-ground inductance, both internal and external to the tube should be kept very low to maintain this advantage. Omission of neutralization will allow a small amount of interaction between the output circuit and the input circuit through the plate-to-filament capacitance. This effect is not very noticeable because of the large coupling between the input and output circuits through the electron beam of the tube. Cathode-driven tetrodes have higher gain and therefore require some form of neutralization.

Grid driven, high gain tetrodes need accurate neutralization for best stability and performance. Self-neutralization can be accomplished very simply by placing a small amount of inductance between the tube screen grid and ground. This inductance is usually in the form of several short, adjustable-length straps. The RF current flowing from plate to screen in the tube also flows through this screen lead inductance. This develops a small RF voltage on the screen, of the opposite phase, which cancels the voltage fed back through the plate-to-grid capacitance. This method of lowering the self-neutralizing frequency of the tube works only if the self-neutralizing frequency of the tube/socket combination is above the desired operating frequency before the inductance is added. Feedback neutralization utilizes a small coupling capacitor, usually in the form of a small plate located near the anode of the tube. The sample of the RF voltage from the anode intercepted by this plate is coupled through a 180 phase shift network into the grid circuit. This technique has the advantage of providing neutralization over a very broad range of frequencies, if implemented correctly and stray reactances are minimized.

Special attention must also be given to minimizing the inductances in the tube socket by integrating distributed bypass capacitors into the socket and cavity deck assembly. Pentodes normally do not require neutralization because the suppressor grid effectively isolates the plate from the grieed. [10]

Power Amplifier Output Circuits

Usually the output circuit consists of a "high-Q" (low loss) transmission line cavity, strip line or a lumped inductor that resonates the tube output capacitance. A means of trimming the tuning and a means of adjusting the coupling to the output transmission line must also be provided by the output circuit. The tank circuit loaded "Q" is kept as low as practical to minimize circuit loss and to maintain as wide an RF bandwidth as possible.

Power Amplifier Cavity

The vacuum tube power amplifier is constructed in an enclosure containing distributed tank circuit elements for minimum loss. The efficiency of the PA depends on the RF plate voltage swing, the plate current conduction angle and the cavity efficiency. The cavity efficiency is related to the ratio of the loaded to unloaded "Q" as follows:

$$N = \left[1 - \left(\frac{Q_L}{Q_U}\right)\right] \times 100 \qquad (8)$$

where:

N = Efficiency in percent
Q_L = Loaded "Q" of cavity
Q_U = Unloaded "Q" of cavity

The loaded "Q" depends on the plate load impedance and output circuit capacitance. Unloaded "Q" depends on the cavity volume and the RF resistivity of the conductors due to skin effects. For best efficiency, a high unloaded "Q" is desirable, as is a low loaded "Q." As the loaded "Q" goes up, the bandwidth decreases. For a given tube output capacitance and power level, loaded "Q" decreases (and bandwidth consequently increases) with decreasing plate voltage or with increasing plate current. The increase in bandwidth at reduced plate voltage occurs because the load resistance is directly related to the RF voltage swing on the tube element. For the same power and efficiency, the bandwidth can also be increased if the output capacitance is reduced. Power tube selection and minimization of stray capacitance are areas of prime concern when designing for maximum bandwidth.

Quarter-Wavelength Cavity

The quarter-wavelength coaxial cavity is the compact and popular PA output circuit illustrated in Figure 4.3-29. The tube anode is coupled through a dc blocking capacitor to a shortened quarter-wave length transmission line. The tube's output capacitance is brought to resonance by the inductive component of the transmission line that is physically less than a quarter-wavelength long. Plate tuning can be accomplished either by adding end-loading capacitance at the high impedance end of the line with a variable capacitor or by changing the position of the ground plane at the low impedance end of the line. The plate tuning capacitor may be a sliding or rotating plate near the anode of the tube. The center conductor of the transmission line (air exhaust chimney) is at dc ground while the anode of the tube operates at a high RF and dc potential. Direct current voltage is fed through an isolated quarter-wavelength de-coupling network inside the chimney to the anode of the tube, while the plate blocking capacitor prevents dc current flow from the anode into the chimney. The quarter-wavelength cavity has approximately twice the operating bandwidth of the half-wavelength cavity.

Folded-Half-Wavelength Cavity

Another approach to VHF power amplification uses the re-entrant, folded-half-wave cavity design illustrated in Figure 4.3-30. The dc anode voltage is applied to the lower portion of the plate line through a choke at the RF voltage null point. The half-wave line is tuned by mechanically expanding or contracting the physical length of a flexible extension (bellows) on the end of the secondary transmission line stub, which is located concentrically within the primary transmission line (air exhaust chimney). Coarse frequency adjustment is accomplished by presetting the depth of the top secondary section of plate line into the tank cavity.

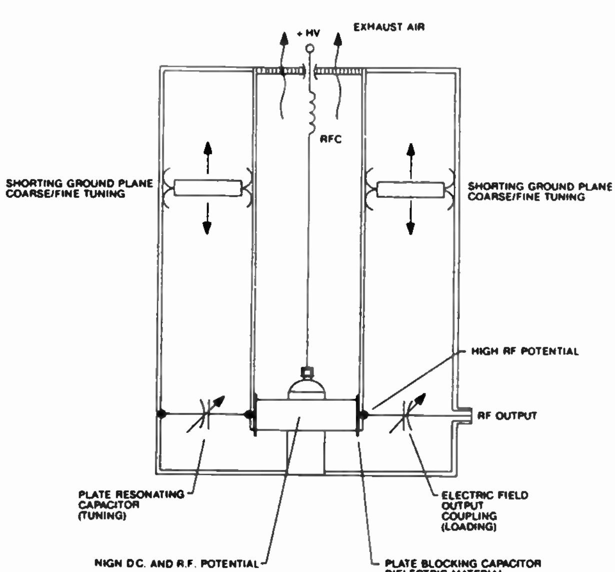


Figure 4.3-29. The quarter-wavelength cavity.

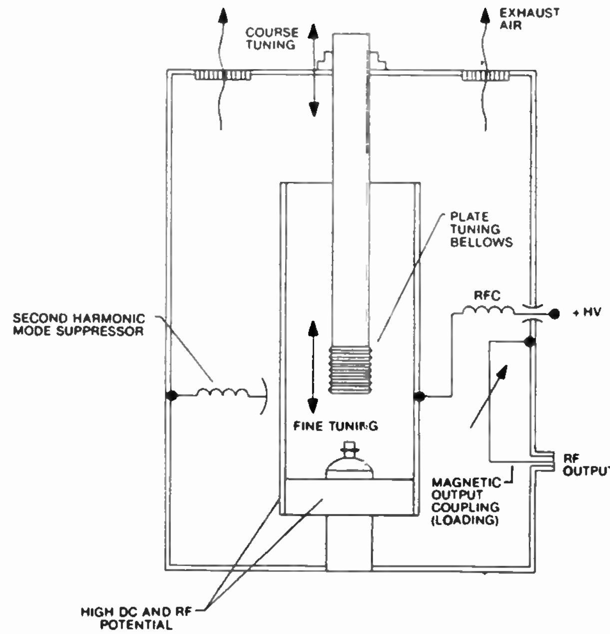


Figure 4.3-30. The folded half wavelength cavity.

Other power amplifier configurations may use lumped components or hybrid combinations with distributed transmission line elements to achieve similar results. The discrete circuit elements are chosen for their individual inductance or capacitance instead of being operated in a purely quarter-wave or half-wave mode. Stray inductance and capacitance add to the component values resulting in the hybrid nature of these circuits.

The RF voltage and current distributions for the quarter-wavelength and the folded-half-wavelength cavities are shown in Figure 4.3-31.

Regardless of the specific configuration, the output circuit must transform the high resonant plate impedance down to the output transmission line impedance of 50 Ω.

The bandwidth of a transmission line cavity is optimized by using a quarter-wavelength and by choosing the highest characteristic impedance mechanically and electrically allowable. [10 and 16] at the end of this chapter give detailed information about the design of tube type RF power amplifiers.

Output Coupling

Power may be coupled from a quarter-wavelength cavity to the transmission line by a capacitive probe located at the high RF voltage point located at the anode end of the quarter-wave line as shown in Figure 4.3-29. The loaded "Q" of this circuit varies with the degree capacitive coupling. Another method of coupling power from the quarter-wavelength cavity uses a tuned loop located near the grounded (high current) end of the line.

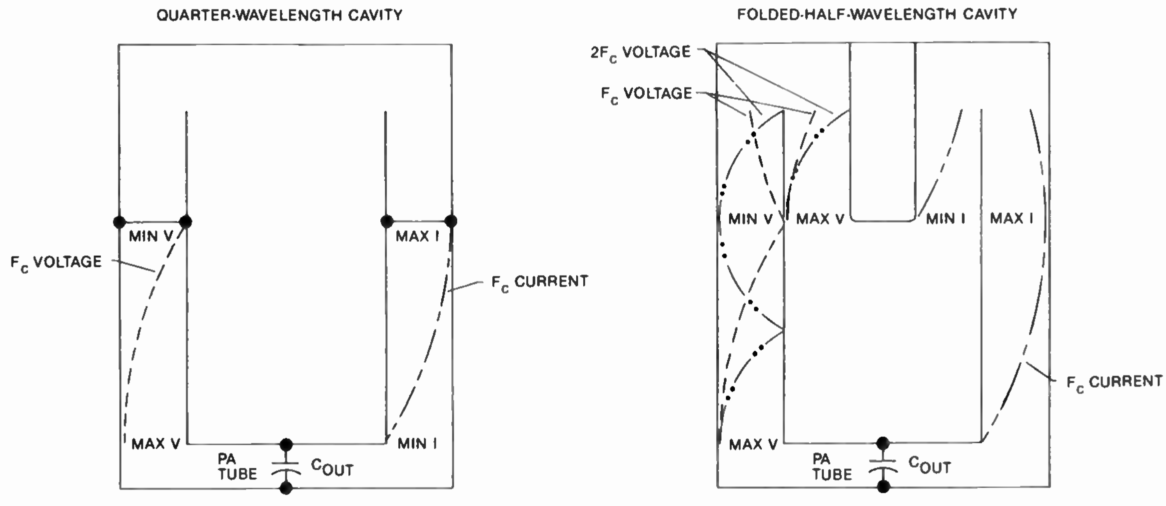


Figure 4.3-31. Cavity RF voltage and current distributions.

In this case, the tuned loop operates both as an inductive and a capacitive pick-up device. Power may be coupled from the half-wave line by an inductive loop located in the strong fundamental magnetic field near the center of the cavity (see Figure 4.3-30).

Power Supplies

Power supplies provide the appropriate dc or ac voltages to the various subsystems with the transmitter. In a typical FM transmitter, the voltages and currents can range from less than 5 V at a few milliamperes to over 10,000 V at several amperes. Safety must therefore be a prime consideration when working around potentially lethal power supplies. Power supplies must be designed with adequate bleeder resistors and interlocks to discharge high voltages before an operator can come in contact with these circuits. The degree to which the ac components are filtered out of the dc outputs of the power supplies will, in large part, determine the "asynchronous" (without FM modulation) AM noise of the FM transmitter.

FM transmitters usually contain multiple power supplies for each of the functional blocks within the system. These power supplies fall into two general categories:

- Single-phase supplies (single input winding on the transformer)
- Polyphase supplies (three or more input windings on the transformer).

Single-Phase Power Supplies

Single-phase power supplies with conventional full-wave rectification and filtering are most often used for the FM exciter, the control system, bias supplies and the IPA. A single-phase supply requires a larger filter choke to achieve the "critical inductance" requirement and a greater value of filter capacitance to maintain acceptably low ripple content compared to a polyphase supply. Large value filter components also mean that the greater stored energy in these components can have a more destructive effect if an arc-over occurs. Choke-input filter sections are normally used to help limit the in-rush current while the shunt capacitor is charging during turn-on. This also maximizes utilization of the transformer and rectifiers by keeping the charging current nearly constant, providing the best filtering action. Choke-input filters have the undesirable characteristic of poor voltage regulation over a wide range of loads. The output voltage will "soar" above the nominal value with no load unless there is enough current through the "bleeder" resistor to keep the choke in the constant current range. Fortunately, in a FM transmitter application, the load on the power supply is fairly constant since the power output of the transmitter does not vary significantly with FM modulation. In higher power transmitters with a three-phase main power source, care must be taken to balance each of the individual single-phase loads among the three phases so that the total load on each of the individual phases is equal.

Polyphase Power Supplies

Polyphase power supplies are used for the final power amplifier, high- voltage supply in high power transmitters. Sometimes they are also used for tube or solid-state IPA supplies. Large blowers used to cool transmitters are usually operated from a three-phase power source. Care must be taken to make the three-line connections to the blower motor in the proper sequence so that the motor will turn in the proper direction.

The most common type of polyphase supply is three-phase with full-wave rectification and LC filtering. Other polyphase systems encountered in broadcast equipment are usually multiples of the three phases with twelve phase rectifiers becoming more popular. The main advantages of a polyphase power supply are:

- Division of the load current between three or more lines to reduce line losses and the size of each of the lines

- Greatly reduced filtering requirements after rectification due to the low ripple at the output of a polyphase full-wave rectifier
- Better voltage regulation with a choke input filter with typically 6% or less "soaring" from no load to full load
- Greater choice of output voltages from a given transformer by selection of either a DELTA or WYE configuration.

The main disadvantage of polyphase systems is their susceptibility to phase imbalance which causes degraded performance of the power supply. If significant imbalance exists in a polyphase system, ripple rejection will be reduced in the polyphase rectifier with a resulting increase in AM noise.

The broadcast engineer should be particularly careful to be sure that the local utility does, in fact, provide true three-phase power to the transmitter site. This can usually be verified by making sure that there are three transformers on the utility pole feeding the transmitter site. In many rural areas, the utilities are still synthesizing pseudo-three-phase service by providing the so-called open-delta (V-V) or Scott (T-T) connection with two transformers instead of true three-phase service. Operation on an open-delta service will degrade the transmitter's performance and increase the susceptibility of the transmitter to damage from transients on the line. Most transmitter manufacturers state that their warranty is void if the transmitter is connected to an open-delta system.

Regulated Power Supplies

In some cases, phase control switching regulation is applied to the high voltage power supply feeding the final output tube. The regulation is accomplished by switching thyristors in the ac mains ahead of the primary winding of the transformer. As the switching duty cycle is reduced, the plate voltage is also reduced. Care must be taken to protect solid-state devices connected to the main power line from transients. Heavy-duty transient suppressors are available for this purpose.

Completely solid-state power amplifiers demand lower voltages at much higher currents. Voltage regulation of these high current supplies is necessary to suppress ripple, but the design of these specialized regulators is different from the typical high voltage power supply. Linear regulators are used at the lower power levels because they are low in efficiency, but they are simple and provide excellent ripple rejection without the need for suppression of switching transients. The linear regulators use series or shunt devices which change resistance dynamically to provide regulation with changes in load and therefore dissipate some of the power within the dynamic resistance.

Switch-mode regulators are used at higher power levels because they are high in efficiency, but they are more complicated and require additional suppression of the switching transients. The high efficiency comes from the digital "on" or "off" nature of the switching regulator which reduces resistive losses by using low loss reactive components to store energy during switching.

Switched tap power supplies can provide good voltage regulation over a limited range at higher efficiency than a switch-mode regulated supply. A solid-state controller switches banks of silicon controlled rectifier (SCRs) between several taps on the secondary of the power transformer to maintain a constant output voltage with changes in load. This type of power supply is simple, efficient and has an excellent power factor (PF). Figure 4.3-32 shows the switching arrangement for the switched tap supply.

Low voltage high current power supplies contain extremely large amounts of stored energy. This can be dangerous due to the high peak currents that can occur during a short circuit across a component with high stored energy. For this reason, special attention must be paid to methods of safely discharging these circuits without damaging components or injuring the operator.

The voltage regulator should provide short-circuit protection with some type of current limiting. The main danger to the operator from this type of power supply is burns due to the nearly instantaneous heating of metallic tools and other conductors (like a metal watchband or ring) that accidentally get into a high current path (such as a short across the filter capacitor).

Power Factor

The PF of a power supply is the measurement of how reactive the input impedance of the power supply appears to the ac power line feeding it and/or how nonlinear this load impedance is. PF is usually specified as a decimal value of less than 1.0 which is the cosine of the phase angle between the voltage and current waveforms on the ac line feeding the input to the power supply.

$$PF = \cos\phi \qquad (9)$$

Where: ϕ is the angle between the voltage and current waveforms

The power utility prefers a resistive load with a unity power factor and minimum disturbance to the voltage waveform on the ac line. A non-unity power factor can be the result of either an inductive or capacitive load. Transmitter power supplies usually have an inductive power factor. Generally, it is desirable to have a power factor of greater than 0.9 which means that the current waveform is displaced less than 26° from the voltage waveform. Minimizing the harmonic content of the waveforms impressed across the ac power line by the power supply is important to prevent interference to other equipment connected to the same power line. If there is high harmonic content in the current waveform, it will degrade the power factor measurement. However, since it is possible with certain linear, reactive loads to have a low power factor with low harmonic content, it is best to measure and specify both the power factor and harmonic content separately from each other. [2, 12, 13]

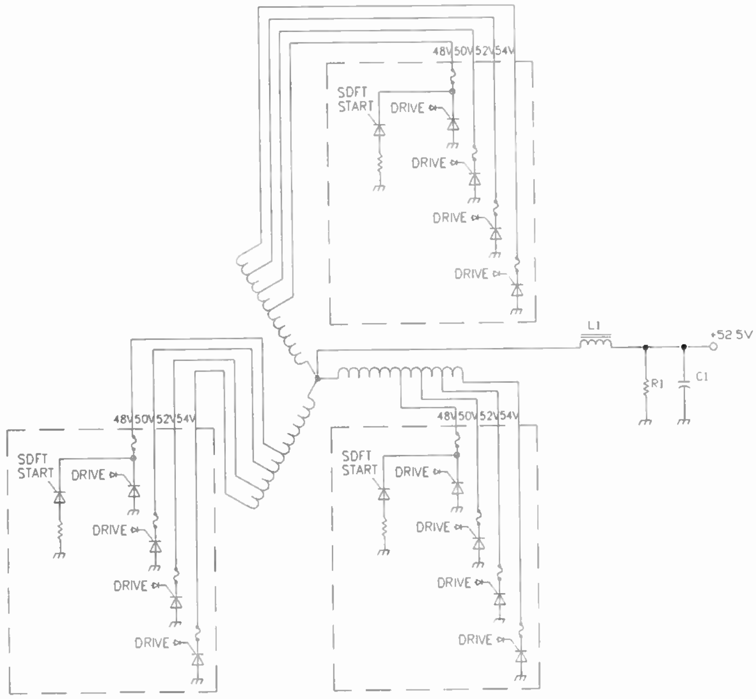


Figure 4.3-32. Switched tap power supply in Harris "Z" solid-state FM transmitter.

Step-Start

Step-start is often used in large power supplies where peak in-rush currents become high when the power supply is first turned on. These peak currents are caused by the need to overcome the hysteresis effect (to initially magnetize the core of the transformer) when ac power is first applied, and to charge the filter capacitor. Step-start systems temporarily insert a resistance or reactance in series with the power lines to limit the current to a reasonable value until initial magnetization of the core and filter charging are completed.

Transmitter Control Systems

Transmitter control systems are often overlooked or given little priority in the selection and maintenance of a broadcast transmitter. The transmitter's control system does, however, serve several important purposes. It provides for:

- Basic on/off control of the transmitter
- Overload protection to protect the transmitter from damage
- Safety interlock protection to prevent injury to people and accessory equipment such as RF switching equipment or RF loads
- A means of controlling the transmitter output power
- Remote control capability and interfacing at installations where the transmitter is not at the same location as the control operator
- Warm-up and cool-down timing sequences of filaments or other time sensitive operations

Additionally, a transmitter control system may provide for:

- Status indications of overloads or other critical parameters
- Automatic regulation of the transmitter output power
- Diagnostic indications to aid in adjustment and maintenance
- Totally automatic operation of the transmitter plant
- Integrated remote control capability.

The transmitter's ability to stay on-the-air will only be as good as the reliability of the control system. So selection and correct operation of the transmitter control system are important.

Early transmitter technology relied on simple relay logic combined with electromechanical contactors to control the transmitter. The speed and variety of overload protection was limited, and diagnostics were not available. While these systems had the virtue of being simple and immune to RFI, the relay logic required continuing maintenance in order to keep the contacts clean and properly adjusted.

Later, transmitter control logic relays were replaced by solid-state digital logic. This provided much more flexibility and reliability, but also raised concerns about the ruggedness and ability of solid-state logic to survive lightning strikes. Although this concern is valid, modern solid-state logic and microprocessor control systems are well protected against damage by optical isolation, shunt protection techniques and radio frequency interference (RFI) filtering. Operating experience with the current generation of transmitters has proven that a properly designed solid-state control system is far more reliable than its relay predecessors.

Automatic Power Control

Many transmitters also provide automatic power control (APC) circuitry to maintain the transmitter's power output within preset limits by correcting for changes in line voltage, component aging or small amounts of drift in operating parameters. The APC circuitry compares a sample of the transmitter output power to a reference and then adjusts the RF drive or other voltages within the transmitter to bring the output power within tolerance.

Some of the more sophisticated APC circuits also provide proportional VSWR fold-back of the transmitter output power. If a sample of the reflected power on the transmission line exceeds a safe limit, the transmitter output power is proportionally reduced to a safe level until the problem is resolved. This feature prevents lost air time during antenna icing or other limited VSWR situations. All standard APC circuits should provide fast-acting shutdown of the transmitter during a catastrophic failure of the antenna system such as a short or open circuit.

Microprocessor Control Systems

Microprocessor technology is now commonly applied to broadcast transmitter control systems. Microprocessor based control systems expand the functions of the transmitter controller from basic housekeeping duties into powerful self-diagnostics, controller redundancy, adaptive control of amplifier combining systems, integrated remote control, user customization and even self-correction of operational problems.

Microprocessor technology lends itself well to industrial control applications like broadcast transmitters. The hardware can be made just as reliable as hard-wired digital logic, allow changes and growth in the operational features by simple modifications in software instructions rather than a complete redesign of the hardware.

Several transmitter manufacturers are marketing transmitters with microprocessor based control systems. Some of the typical features that distinguish these control systems from non microprocessor systems are:

- Built-in "trouble tree" with fault location and diagnostic read-outs with user-friendly messages
- Simultaneous read-outs of all operating parameters
- Real time calculation of efficiency, dissipation, VSWR and other parameters requiring mathematical operations
- Adjustment of parameters to maximize efficiency and minimize reject load dissipation
- Built-in clock/calendar for logging changes in operating status, power failures and overloads
- Tolerance flagging on key operating parameters as warnings for logging and for preventive maintenance
- Ability to communicate with the outside world for remote control or logging purposes through a standard serial interface or TCP/IP
- Integrated remote control capability without external remote control equipment and interfacing
- Provision to customize the system features to the station's individual requirements through the use of software menus
- Tuning aids which allow the operator to adjust the system for peak efficiency, minimum dissipation and minimum VSWR by means of a real time display of these calculated parameters

The method of communicating information to the operator varies among systems, but most use LED or LCD read-outs with codes or alphanumeric messages. Microprocessor controllers can also be equipped with large graphic displays so that a large amount of information can be displayed in several different graphical formats.

Controller Back-Up Systems

A certain degree of redundancy is desirable in the transmitter control system so that the transmitter can stay on-the-air even if a portion of the system fails. There are several approaches currently in use to provide back-up systems. A multi-level hierarchy can be used which automatically hands over basic control functions to a primary "life support" controller in the event of microprocessor hardware or software problems. Good systems design separates diagnostic and supervisory functions from basic control functions so that a failure in a higher level function will not affect the ability of the system to remain on-the-air without interruption. Watchdog circuits and software are embedded within the control system to detect failures and initiate corrective action before an interruption in service occurs.

It is also possible to have distributed microprocessor systems with multiple processors that can automatically pick up the tasks of a failed processor without affecting the ability of the transmitter to remain on the air. The ability to quickly replace a controller subsystem while remaining on the air is an advantage of the latest generation of transmitter control systems.

Remote Control Interfacing

Regardless of the type of control system used, the ability to interface easily with standard remote control systems is very important. Most modern systems interface with parallel control lines for each individual function requiring a momentary contact closure of 24 V or less at a current of 50 milliamperes or less. These levels are compatible with relay logic or optically isolated solid-state logic. Analog levels output from the transmitter for remote meter readings generally are fully buffered and fall into the range of 0–5 V dc for a full scale reading at an impedance level of less than 10 KΩ.

The advent of microprocessor based control systems has been accompanied by a trend toward using standard computer asynchronous serial interfacing instead of parallel interfacing. Serial interfacing reduces the number of connections to the transmitter and can carry both control functions and digitized meter readings through the same interface. By converting analog information into digital information before transmission to the remote control point, the need for calibration and recalibration of the remote metering point is reduced. The current trend is toward open, nonproprietary, serial data protocols and interface standards such as Internet SNMP and TCP/IP. This allows the user maximum flexibility in choosing dedicated remote control equipment or using software based network management tools. Microprocessor based control systems also allow the remote control system to have access to more in-depth information about the transmitter than is possible by parallel interfacing with an external remote control system. A personal computer or laptop computer can be used to control the transmitter through an ordinary dial-up phone line, Internet IP address or radio link.

For more specific information, see Chapter 4.7 "Transmitter Remote Control Systems."

RF Output and Systems Filters

The high efficiency, nonlinear RF power amplifiers used in FM broadcast transmitters generate significant amounts of energy on frequencies that are integer multiples (harmonics) of the desired fundamental frequency. The output circuit alone does not provide enough harmonic attenuation to meet FCC regulations. To comply with Part 73 of the CFR and to prevent interference to other services, a low-pass filter must be installed in the transmission line at the output of the transmitter. The FM band is narrow enough that one low-pass filter design can be used for any FM channel carrier frequency. These filters usually consist of multiple LC sections arranged so that frequencies within the FM band are passed with little attenuation (typically 0.1 dB or less) while frequencies above the FM band are highly attenuated (60 dB or more).

The most common type of filter used in this application is a reflective filter. This means that the frequency components outside the passband are reflected back out of the filter toward the source because the filter provides a mismatch at these undesired frequencies. The filter can be constructed using either lumped inductors and capacitors or by using a section of nonconstant impedance transmission line to form distributed inductors and capacitors. The filters designed for low power transmitters often employ "lumped" elements (coils and capacitors) because these elements are compact and can be integrated into the transmitter cabinet. The distributed type of filter is most often used with high power-FM broadcast transmitters because of its simplicity, extreme ruggedness and ability to handle higher power levels. The distributed filter does have the disadvantage of having larger physical dimensions than a similar lumped filter, which may necessitate mounting the filter external to the transmitter cabinet. Figure 4.3-33 shows a cut-away view of a typical distributed low-pass filter. Note that the areas where the center conductor of the transmission line is smaller than that required for the input Z_0 are inductive, while the areas where the center conductor is larger in diameter are capacitive.

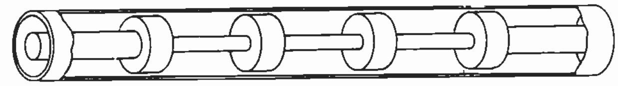


Figure 4.3-33. Cut-away view of a distributed low pass filter.

When two filters (such as the output cavity and the harmonic filter) are connected together by a transmission line, the total harmonic attenuation will vary with interconnecting line length. The attenuation characteristics of the harmonic filter are specified for the condition where both the source and load impedances are equal to the desired transmission line impedance.

In actual use, the source impedance at the output of the tank circuit is much less than the 50 Ω load impedance presented by a properly terminated filter. If an incorrect length of line is selected, the harmonic attenuation may be insufficient and the transmitter tuning may be affected. This undesirable condition can be corrected by changing the line length by approximately one-quarter wavelength. The line length between the tank circuit and harmonic filter is usually supplied precut to a value satisfactory to the transmitter manufacturer.

Harmonic Notch Filters

In some cases, a second harmonic notch filter is required in addition to the low-pass filter because the second harmonic component from the amplifier is high and the cut-off slope of the low-pass filter is not steep enough to provide sufficient second harmonic attenuation. The additional attenuation required (typically 30 dB) can be provided by a notch filter which places a short circuit across the transmission line at the second harmonic while providing a high impedance at the fundamental. A one-quarter wavelength (at the fundamental frequency) shorted coaxial stub is often used for this function. The second harmonic energy is partially reflected back into the power amplifier and partially dissipated in the equivalent series resistance of the series-tuned circuit formed by the stub. This shorted stub provides a very low inductance and a dc path from the center conductor of the transmission line to ground, providing a separate, protective advantage by shunting static discharges, such as lightning, to ground. Some transmitters have internal second harmonic suppressors which eliminate the need for an external notch filter.

Transmission Line Power and SWR Measurements

Directional wattmeters are instruments that measure the forward power (P_f) and reflected power (P_r) in a transmission line. The net power delivered to the load (antenna) is ($P_f - P_r$). If the transmission line is perfectly matched, all the forward power will be absorbed by the load and there will be no reflected power. The peak voltage at each point along the line will be the same value; similarly, the current at each point along the line will also have a uniform value. If the transmission line is mismatched, there will be reflected power with a resulting standing wave on the line. This means that the voltage and current distributions along the line will no longer be uniform with high values at certain points on the line and low values at points one-quarter wavelength away. The ratio of high value to the low value is called the standing wave ratio (SWR).

SWR

The SWR on the transmission line is related to the ratio of the forward to reflected power by the following formula:

$$\mathrm{SWR} = \frac{1 + \sqrt{(P_r/P_f)}}{1 - \sqrt{(P_r/P_f)}} \qquad (10)$$

This relationship is shown graphically in Figure 4.3-34 so the approximate SWR can be obtained without computation.

The standing wave is due to the presence of two components of power, one traveling toward the load and the other reflected by the load mismatch, traveling back toward the generator.

These components are defined as:

$$P_f = \frac{E_f^2}{Z_o} = I_f^2(Z_o) \qquad (11)$$

$$P_r = \frac{E_r^2}{Z_o} = I_r^2(Z_o) \qquad (12)$$

$$P_n = (P_f - P_r) \qquad (13)$$

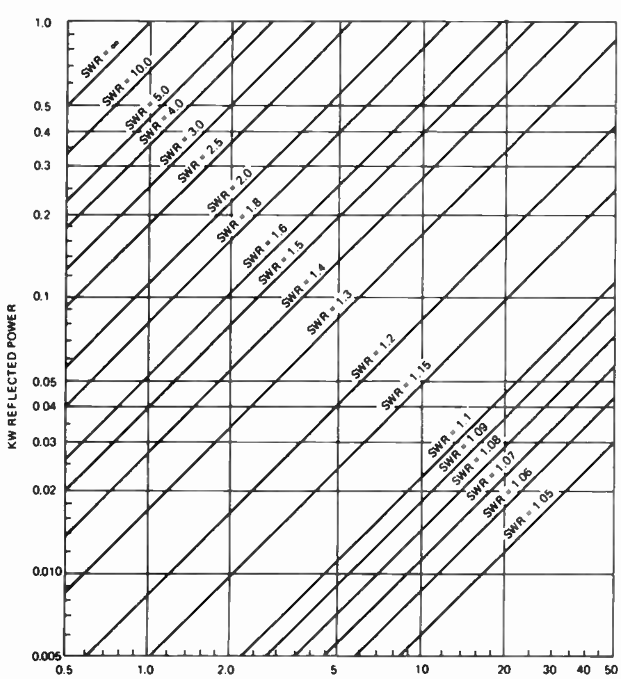


Figure 4.3-34. Chart of SWR versus forward and reflected power.

The subscripts ($_f$) and ($_r$) are used to denote the forward and reflected values of power, voltage and current; while Z_o is the characteristic impedance of the transmission line. (P_n) is the net power absorbed by the load (transmission line loss and antenna radiation).

Since the forward and reflected voltage and currents are traveling in opposite directions, they will add in phase at some point along the line length to produce a voltage maximum. One-quarter wave length along the line in either direction from this maximum, the forward and reflected components are out of phase and produce a voltage minimum. The forward and reflected components of current also add vectorially to produce a current standing wave. The magnitude of the standing wave is defined as:

$$\mathrm{SWR} = \frac{E_{\max}}{E_{\min}} = \frac{I_{\max}}{I_{\min}} \qquad (14)$$

The ratio of the highest voltage point on the line to the lowest voltage point on the line is a commonly used measure of system performance defined as the VSWR. A VSWR of 1.0:1 means that a perfect match has been achieved, while a VSWR of 2.0:1 means that a mismatch is causing approximately 11% of the power to be reflected.

At the point of reflection (the load mismatch), the phase of the reflected current is reversed 180 from the forward current. The reflected voltage does not have this phase reversal. This displaces the voltage and current standing waves by 90° along the line so that the ($E_{\max}$) and ($I_{\min}$) occur at the same points while ($E_{\min}$) and ($I_{\max}$) occur 90° (one-quarter wavelength) away, in either direction from ($E_{\max}$) and ($I_{\min}$).

Since the reflected current is reversed in phase a directional coupler can measure forward and reflected power separately. A small voltage is obtained by inductive coupling which represents the current in the transmission line. To this is added a sample of the voltage across the line that is simultaneously obtained by capacitive coupling. These two samples are adjusted to be exactly equal when the line is terminated with its characteristic impedance (no standing waves and no reflected components). The two RF samples are added, which gives a resultant RF voltage proportional to the forward components of voltage and current as illustrated in Figure 4.3-35(a).

The forward components of the samples are equal and in phase, but the reflected components of voltage and current balance out. By having a second coupling section physically turned around in the opposite direction, the phase of the current sample is reversed; the reflected components add while the forward components balance out as illustrated by the vector diagram in Figure 4.3-35(b). The voltages representing the forward and reflected powers output by the directional coupler are usually rectified and buffered to feed the automatic power controller and a power indicating meter. Since power is proportional to the square of the voltage on the transmission line, the meter scale is calibrated to read the square of its input. It also has

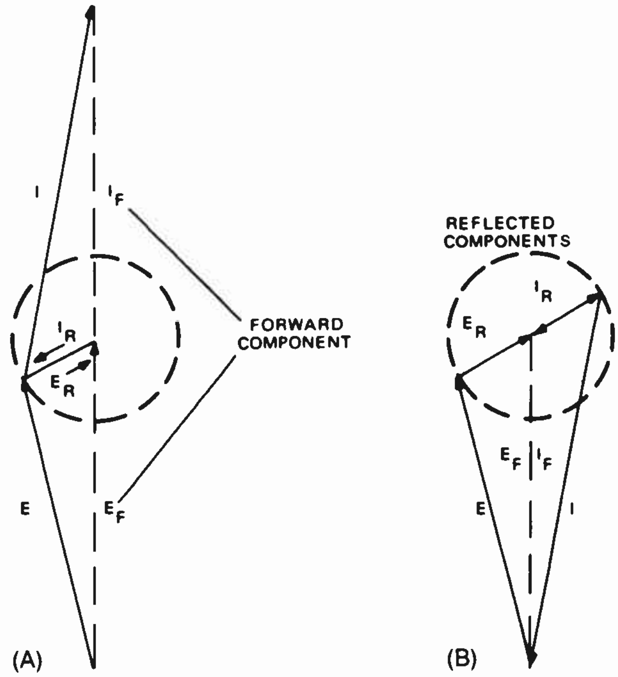


Figure 4.3-35. Phasor addition of voltage and current samples to separate forward and reflected components.

a diode correction factor so that forward and reflected power are read out directly.

VSWR Measurement

Although some FM transmitters can operate into a VSWR of greater than 1.8:1, the VSWR on an FM antenna transmission line should normally be kept to a value of 1.1:1 or less for good stereo performance. It takes very little reflected power to produce substantial VSWR (see Figure 4.3-34). For this reason, the reflected power is usually read on a more sensitive meter position. Problems in the antenna system, such as loose connections or icing, may cause excessive VSWR. Instruments external to the transmitter are available that monitor reflected power and energize an alarm if it becomes excessive. As long as the transmitter power output is fairly constant, the use of reflected power to indicate excessive VSWR is simple and adequate.

Combined Transmitters

It is possible to combine the output of two RF power amplifiers for higher power levels. The important advantage is that the broadcast transmission is not interrupted if one amplifier fails. The radiated signal strength merely drops 6 dB until the failed amplifier is repaired and put back on the air. A dual amplifier system costs more than a single amplifier for a given total power output, but there are the economic advantages of reducing lost air time and eliminating the need for a separate standby transmitter. Automatic or manual output switching can be used to route the full power of the remaining amplifier directly to the antenna.

Two methods may be used to bypass the output combining hybrid to allow 100% of the power of the remaining transmitter to be sent to the antenna if one transmitter of a combined pair should fail.

The first method uses three motorized switches (or patch panels) to bypass the 3 dB hybrid while connecting the operating transmitter directly to the antenna and the failed transmitter directly to the test load. This allows recovery of the 50% power lost in the reject load when one transmitter is off the air. One disadvantage is that the system must be taken off the air for several seconds for operation of the coax switches.

A second method provided by some transmitter suppliers uses a pair of 3 dB hybrids interconnected with one fixed and one variable RF phasing section. The phasing section is constructed to operate while under RF power and can re-direct the full output of either transmitter directly to the antenna and place the other transmitter into the test load without taking the system off the air. A dedicated system controller allows automatic or manual control. This so-called switchless combiner offers the highest possible on-air availability for combined FM transmitters. With complete redundancy in the RF power amplifier chain, some stations go one step further and also install dual exciters with automatic switching so that if one exciter fails, the other unit is quickly switched into service.

Hybrid Couplers

Hybrid couplers are reciprocal four-port devices that can be used either for splitting or combining RF sources over a wide frequency range. Figure 4.3-36 shows an exploded view of a typical 3 dB, 90°, hybrid coupler. The coupler consists of two identical parallel transmission lines that are coupled over a distance of approximately one-quarter wavelength and are enclosed within a single outer conductor. Ports at the same end of the coupler are in-phase while ports at opposite ends of the coupler are in quadrature (90° phase shift) with respect to each other.

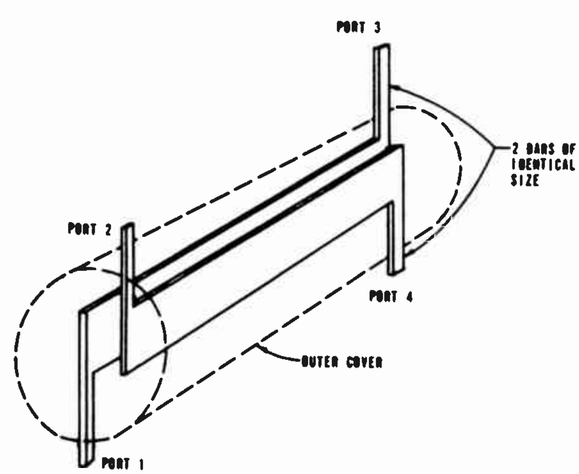


Figure 4.3-36. Physical model of 90° hybrid coupler.

The phase-shift between the two inputs or outputs is always 90° and is almost independent of frequency. If the coupler is being used to combine two signals into one output, these two signals must be fed to the hybrid coupler in phase quadrature. The reason this type of coupler is also called a 3 dB coupler is that when used as a power splitter, the split is equal or half-power (3 dB) between the two outputs.

Hybrid Combiners

The output hybrid combiner effectively isolates the two amplifiers from each other. Tuning adjustments can be made on one amplifier, including turning it on and off, without appreciably affecting the operation of the other amplifier. Good isolation is necessary so that if one transmitter fails, the other will continue to operate normally instead of in a mistuned condition. Two of the ports on the hybrid coupler are the inputs from the power amplifiers. The sum port is the antenna output terminal and the difference port goes to a resistive dummy load called the "reject load" since only the rejected power due to imbalance appears here. When the power fed to each of the two inputs is equal in amplitude with a phase difference of 90° the total power is delivered to the sum port (antenna). Very little of the power appears at the reject load if the phase relationship and power balance are correct. If the phase relationship is reversed between the two amplifiers, all the power is delivered to the reject load, so care must be taken to be sure that the proper one of the two possible 90° phase relationships is used. When all the ports on the hybrid combiner are properly terminated, isolation of 30 dB or more can be achieved between the power amplifiers. For perfect isolation between the amplifiers, the load impedance on the sum and difference ports must be exactly the same. This is approached in practice by providing a 1.0:1 VSWR with a resistive 50 Ω load for the termination (reject load) on the difference port and then reducing the VSWR on the antenna transmission line as low as possible by trimming the antenna match. This will keep the input port impedances from changing very much when one amplifier is not operating.

The input ports will present a load to each transmitter with a VSWR that is lower than the VSWR on the output transmission line. This is because part of the reflected power coming into the output port will be directed to the reject load and only a portion will be fed back into the transmitters. Figure 4.3-37 shows the effect of output port VSWR on the input port VSWR and on the isolation between ports.

If the two inputs from the separate amplifiers are not equal in amplitude or exactly in phase quadrature, some of the power will be dissipated in the difference port reject load. The match in input power and phase is not extremely critical. See Figures 4.3-38 and 4.3-39.

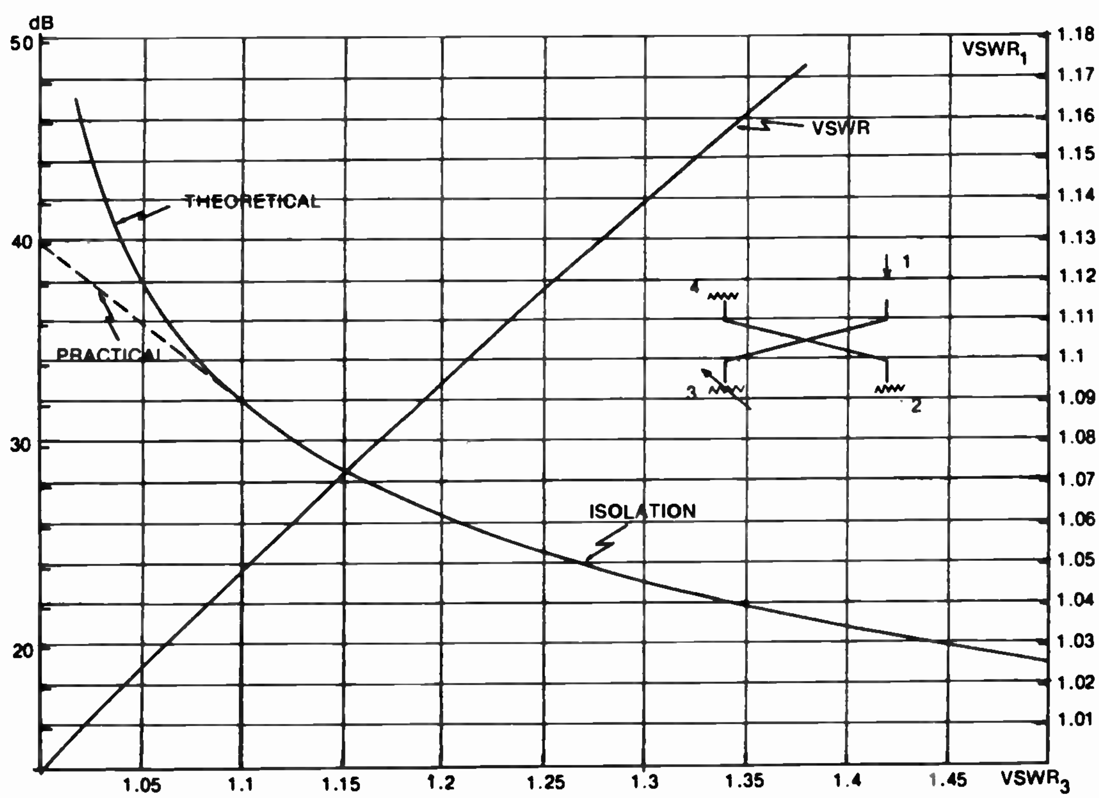


Figure 4.3-37. Isolation of hybrid coupler.

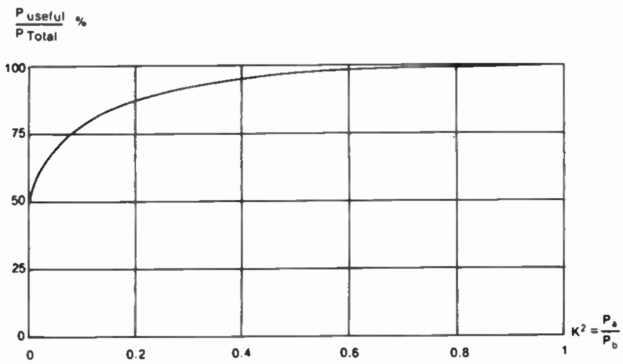


Figure 4.3-38. Power imbalance in hybrid couplers.

The power lost in the difference port reject load can be easily reduced to a negligible value by touching up the amplifier tuning and by adjusting the phase shift. For example, if one amplifier is delivering only half the power of the other amplifier, only about 3% of the total available power will be dissipated in the reject load; 97% is still fed to the output transmission line.

If one transmitter fails completely, half of the working amplifier's output goes to the antenna, and the other half is dissipated in the difference port reject load. This is why the radiated output drops by 6 dB or to one-fourth of the original combined power. The reject load must be rated to handle a minimum of one-fourth of the total combined power, but often the reject load is rated to handle one-half the total power so that it can also be used as a test load for one of the transmitters.

Hybrid Splitting of Exciter Power

Figure 4.3-40 shows a block diagram of a pair of combined amplifiers with dual exciters. The exciters cannot be operated in parallel like the amplifiers because their RF outputs would have to be on exactly the same carrier frequency and almost exactly in phase under all modulation conditions. An automatic or manual exciter switcher is used to direct the output of the

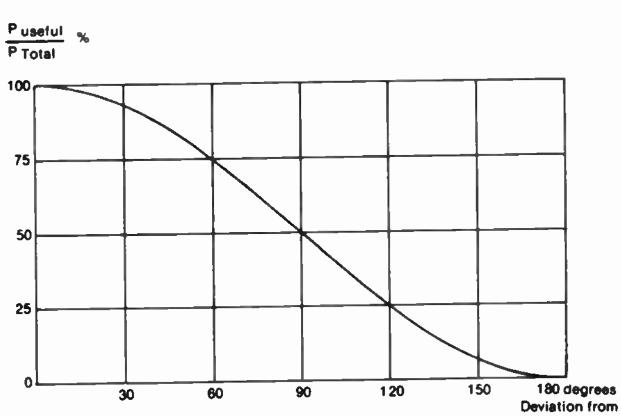


Figure 4.3-39. Phase sensitivity, hybrid coupler.

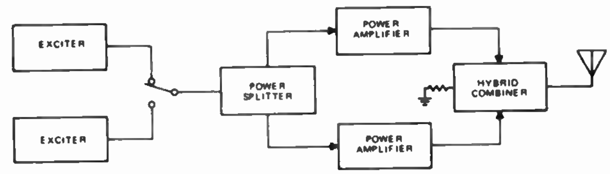


Figure 4.3-40. Block diagram of transmitter with two power amplifiers, a hybrid combiner and dual exciters.

desired exciter to the combined transmitter while the other stand-by exciter is routed to a dummy load. The one exciter in use feeds a hybrid splitter/phase shifter which transforms one 50 Ω input into two isolated 50 Ω outputs that have a 90° phase shift between them with half the power going to each output. The operation of this hybrid splitter is the reciprocal of the hybrid combiner described above. The exciter must have enough power output capability to drive both power amplifiers. In some cases, an additional IPA is required between the exciter and the splitter to boost the drive level. The length of coax from the power splitter to each amplifier input must be cut to a precise length so that the amplifiers will be fed in the proper phase relationship.

Each of the power amplifiers is assumed to have equal gain and phase shift. In practice, it may be difficult to get the amplifiers tuned so that their gains and phase shifts are equal at the same time. For this reason, a line stretcher or variable phase shift network is usually included with the exciter splitter so that the station engineer will have the ability to adjust phasing independent of amplifier tuning. For more detailed information on the theory of hybrid couplers, see the chapter on television transmitters and [7].

Filterplexing

Having several FM stations share a single broad band antenna system is becoming more and more popular. A special device called a filterplexer (also known as an RF multiplexer) is used to connect several transmitters on different frequencies together onto one antenna system. The filterplexer provides isolation between the various transmitters while efficiently combining their power into a single transmission line. This is usually accomplished by a system of band-pass filters, band-reject filters and hybrid combiners. The isolation is required to prevent power from one transmitter from entering another transmitter with resulting spurious emissions as well as to keep the rest of the system running in the event of the failure of one or more transmitters.

An important consideration in the design of a filterplexing system is the effect on the phase response (group delay characteristic in the passband) of each of the signals passing through the system due to individual bandwidth limitations on each of the inputs.

RF Intermodulation Between FM Broadcast Transmitters

Interference to other stations within the FM broadcast band, as well as to other services outside the broadcast band, can be caused by RF intermodulation

3rd ORDER INTERMODULATION PRODUCTS

f_1=100.3 MHz. f_2=101.1 MHz.

$2f_1$-f_2=[2(100.3)-(101.1)]=[200.6-101.1]=99.5 MHz.

$2f_2$-f_1=[2(101.1)-(100.3)]=[202.2-100.3]=101.9 MHz.

OR

[f_1-(f_2-f_1)]=[100.3-(101.1-100.3)]=[100.3-0.8]= 99.5 MHz.

[f_2+(f_2-f_1)]=[101.1+(101.1-100.3)]=[101.1+0.8]=101.9 MHz.

4.3-41. Example of how the intermodulation product frequencies may be calculated.

between two or more FM broadcast transmitters. Transmitter manufacturers have begun to characterize the susceptibility of their equipment to RF intermodulation so this information is becoming available to the designers of filterplexing equipment.

The degree of intermodulation interference generated within a given system can be accurately predicted before the system is built if the actual mixing loss of the transmitters is available when the system is designed. Accurate data on *"Mixing Loss" or "Turn-Around-Loss"* not only speeds the design of filterplexing equipment, but also results in higher performance and more cost-effective designs because the exact degree of isolation required is known before the system is designed. Filterplexer characteristics, as well as antenna isolation requirements, can be tailored to the specific requirements of the transmitters being used. The end user is assured in advance of construction that the system will perform to specification without fear of overdesign or underdesign of the components within the system.

Mechanisms that Generate RF Intermodulation Products

When two or more transmitters are coupled to each other, new spectral components are produced by mixing the fundamental and harmonic terms of each of the desired output frequencies. For example, if only two transmitters are involved, the third order intermodulation (IM_3) terms could be generated in the following way. The output of the first transmitter (f_1) is coupled into the nonlinear output stage of the second transmitter (f_2) because there is not complete isolation between the two output stages. The (f_1) will mix with the second harmonic of (f_2), producing an in-band third order term with a frequency of $[2(f_2) - (f_1)]$. In a similar fashion the other third order term will be produced at a frequency of $[2(f_1) - (f_2)]$. This implies that the second harmonic content within each transmitter's output stage, along with the specific nonlinear characteristics of the output stage, will have an effect on the value of the mixing loss.

It is possible, however, to generate these same third order terms in another way. If the difference frequency between the two transmitters $[(f_2) - (f_1)]$, which is an out-of-band frequency, re-mixes with either (f_1) or (f_2), the same third order intermodulation frequencies are produced.

Empirical measurements indicate that the $[2(f_2) - (f_1)]$ type of mechanism is the dominant mode generating third order IM products in modern transmitters using a tuned cavity for the output network.

Figure 4.3-41 shows an example of how the intermodulation product frequencies may be calculated. Figures 4.3-42 and 4.3-43 show the resulting frequency spectra.

Intermodulation as a Function of "Turn-Around Loss"

Turn around loss or mixing loss describes the phenomenon whereby the interfering signal mixes with the fundamental and its harmonics within the nonlinear output device. This mixing occurs with a net conversion loss; hence, the term turn around loss has become widely used to quantify the ratio of the interfering level to the resulting IM_3 level. A turn around loss of

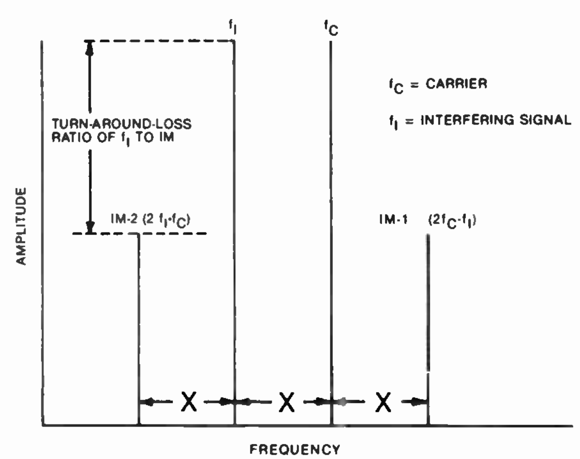


Figure 4.3-42. Frequency spectrum of third order IM with the interfering signal level equal to the carrier level.

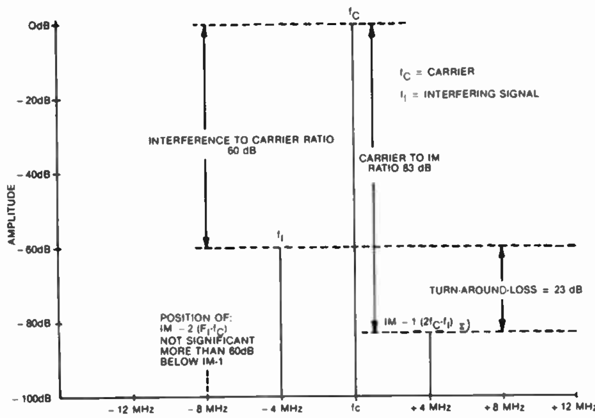


Figure 4.3-43. Typical frequency spectrum of third order IM of a broadcast FM transmitter.

10 dB means that the IM_3 product fed back to the antenna system will be 10 dB below the interfering signal fed into the transmitter's output stage.

Turn around loss will increase if the interfering signal falls outside the passband of the transmitter's output circuit, varying with the frequency separation of the desired signal and the interfering signal. This is because the interfering signal is first attenuated by the selectivity going into the nonlinear device; the IM_3 product is then further attenuated as it comes back out through the frequency selective circuit.

Turn around loss consists of three components:

1. The basic in-band conversion loss of the nonlinear device
2. The attenuation of the out-of-band interfering signal due to the selectivity of the output stage
3. The attenuation of the resulting out-of-band IM_3 products due to the selectivity of the output stage.

As the turn around loss increases, the level of undesirable intermodulation products is reduced, and the amount of isolation required between transmitters is also reduced.

The transmitter output circuit loading control directly affects the source impedance. This therefore affects the efficiency of coupling the interfering signal into the output circuit where it mixes with the other frequencies present to produce IM_3 products. Light loading reduces the amount of interference that enters the output circuit with a resulting increase in turn around loss. In addition, the output loading control setting will change the output circuit bandwidth (loaded "Q") and therefore will also affect the amount of attenuation that out-of-band signals will encounter passing into and out of the output circuit.

Second harmonic traps or low-pass filters in the transmission line of either transmitter have little effect on the generation of intermodulation products. This is because the harmonic content of the interfering signal entering the output circuit of the transmitter has much less effect on IM_3 generation than the harmonic content within the nonlinear device itself. The resulting IM_3 products fall within the passband of the low-pass filters and outside the reject band of the second harmonic traps, so these devices offer no attenuation to RF intermodulation products.

Figure 4.3-44 gives an overview of the various filtering options to prevent excessive IM_3 products.

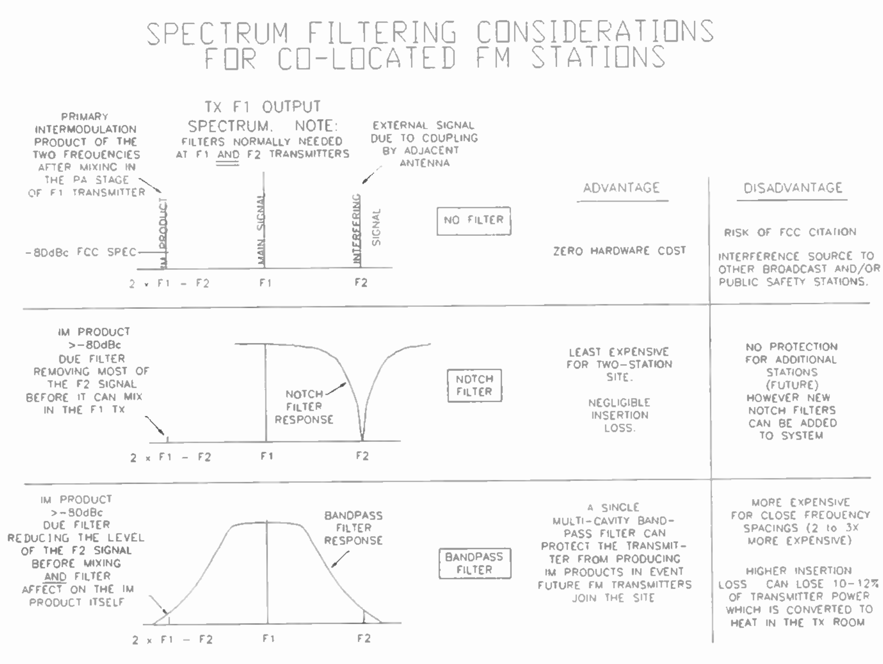


Figure 4.3-44. RF filtering requirements.

Reduction of RF Intermodulation in Solid-State Transmitters

Depending on the topology of the combining system, broad band, solid-state transmitters are more likely to generate RF intermodulation products than single-tube transmitters which have a narrow band, selective cavity in the output stage. However, some modern solid-state designs use balanced "N-way" module combiners together with a conventional 3 dB hybrid as the final output combining stage. This topology tends to provide a rather uniform 15 dB turnaround loss to incoming RF interference, regardless of the frequency separation of the solid-state transmitter's FM carrier from an interfering carrier. For frequency spacings closer than about 5 MHz from the interfering carrier, a solid-state transmitter using this combining technique actually produces fewer RF intermodulation IM_3 products than a single tube FM transmitter whose close in turnaround loss can be less than 6 dB. In some cases, a solid-state FM transmitter of this design may not require as much (or any) external RF filtering. In other cases, it may require more RF filtering than a tube/cavity transmitter. The turnaround loss of a narrow band tube/cavity transmitter is usually better than a broad band, solid-state transmitter when there is a large frequency separation (>5 MHz) between the carrier frequency and the interfering frequency. This is due to the skirt selectivity of the narrowband final amplifier cavity. Co-sited FM transmitters are likely to need external RF filtering to prevent the generation or transmission of unacceptable RF intermodulation products, regardless of whether they are of tube/cavity or of solid-state design. For more detailed information about RF intermodulation between transmitters, see the chapters on FM and TV antenna systems and [14].

Synchronous FM Boosters

According to FCC rules, FM stations may operate on frequency booster facilities up to 20% of the maximum permissible ERP for the class of primary station they rebroadcast. The booster station may not, however, transmit beyond the predicted 1 mv/m contour of the main transmitter for Class A and Class C stations, the 0.5 mV/m contour for Class B, or the 0.7 mV/m contour for Class B1 stations.

Effects of Adding a Booster

For the purpose of analyzing the effects of a second carrier, the addition of a booster signal can be treated as interference. A second interfering carrier will both amplitude and phase (frequency) modulate an existing desired carrier. The characteristics of this apparent modulation are given by Equation 15.

$$F_m = |f_c - f_i| \qquad B = A_i/A_c \tag{15}$$

Where:

F_m = Frequency separation of the carriers
f_c = Main carrier frequency
f_i = Booster (interfering) carrier frequency
A_i = Booster (interfering) carrier amplitude
A_c = Main carrier amplitude
B = Percentage of amplitude modulation

Equation 15: Characteristics of an Interfering Signal Modulating a Desired Signal.

In other words, an FM receiver detecting two carriers (unmodulated for simplicity), decodes a modulation tone equal in frequency to the absolute value of the frequency separation between the carriers. Moreover, the modulation index (both AM and FM) is simply the ratio of the carrier amplitudes. Note, however, that the modulation index is never more than one, since increasing the amplitude of the interfering signal over that of the original carrier simply makes the carrier the interfering signal to the booster. For FM, B is measured in radians (see Equation 15), while for AM, B is the percentage of amplitude modulation produced.

Synchronous Carriers

This gives necessitates the need for synchronizing carrier frequencies. An analysis of the equation for B (FM) shows, given a fixed carrier ratio (fixed modulation index), an increase in carrier frequency separation is equivalent to an increase in Δf. For FM this is equivalent to an increase in detected signal amplitude (see Equation 16).

If, given $B = \Delta f/F_m$, B is held constant and F_m is the frequency separation of the carriers, then:

$$\Delta f = B \times F_m \tag{16}$$

Equation 16: FM modulation index, alternate form. This phenomenon is shown graphically in Figure 4.3-45.

Figure 4.3-45 illustrates the relationship of carrier frequency separation to equivalent carrier frequency deviation (B = 1).

Knowing this, the advantages of frequency locking the carriers becomes obvious. By taking the limit as the difference in carrier frequency approaches zero, two things happen. First, the frequency of the detected tone approaches zero and the equivalent FM deviation

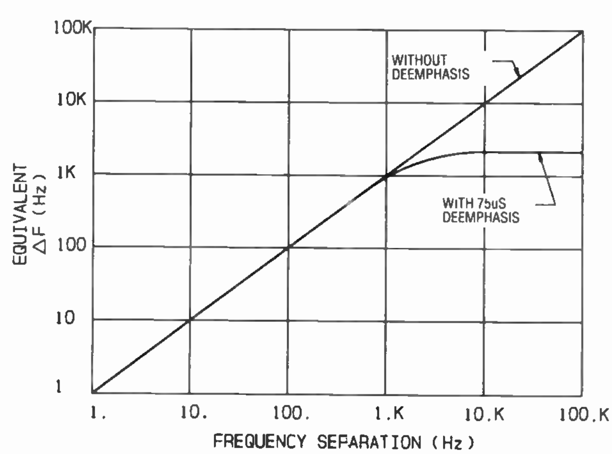


Figure 4.3-45. Relationship of carrier frequency separation to equivalent carrier frequency deviation (B = 1).

produced by that tone approaches zero. In other words, the interference disappears.

The Resultant Carrier

This leaves us with a single frequency carrier whose amplitude depends on the relative phase relationship between the two signals at any given reception point. If the signals are in phase and the carrier ratio is one (0 dB), then the resultant amplitude is twice that of either carrier. On the other hand, if the carriers are 180° (out of phase), the net result is zero. Between these two extremes, the resultant amplitude and phase are given by Equation 17 and 18 respectively.

$$A_r = \sqrt{[A_c + A_i \text{ Cos } \omega_i(t))^2 + (A_i \text{ Sin } \omega_i(t))^2]} \tag{17}$$

where:

A_r = Resultant Carrier Amplitude
A_c = Main Carrier Amplitude
A_i = Booster Carrier Amplitude
$\omega_i(t)$ = Angle between A_i and A_c

and:

$$\Phi_r = \text{Arctan} \frac{A_i \text{ Sin } \omega_i(t)}{A_r + A_i \text{ Cos } \omega_i(t)} \tag{18}$$

where: Φ_r = Resultant Carrier Phase Angle

Figure 4.3-46 shows this vector summation.

Resultant Effect on Receivers

All the conditions for adequate reception would be met if we were in a reception area of equal in phase and frequency locked carriers, located on a straight line between the two transmitting antennas and the information on each carrier were occurring at the same time.

However, if we were to move 2.5 ft toward either transmitter (roughly equivalent to one-quarter wavelength at 100 MHz), we would now be in an area where the two signals are 180° out of phase and there is no longer a signal to detect. Notice that we have only moved one-quarter wavelength, yet have actually moved into a null! This is because we have moved one-quarter wavelength away from one transmitter and one-quarter wavelength closer to the other, giving a total one-half wavelength change. In a mobile receiver,

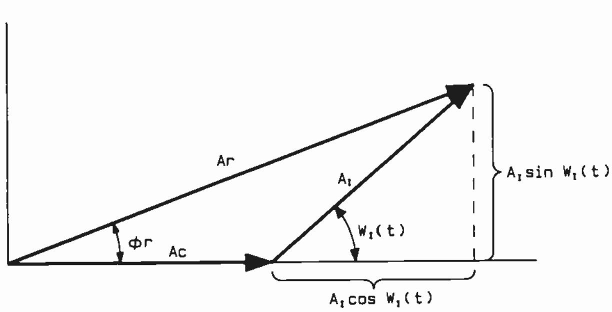


Figure 4.3-46. Vector summation of two carriers.

this phenomenon is virtually identical to the "picket fencing" of multipath, except that the interfering signal is not reflected but rather is a duplicate transmission from a booster site.

For the stationary receiver in an additive state, this is not a problem, and, in fact, gross modulation problems can be effectively covered up by virtue of the capture effect of an FM receiver, provided the carrier ratios are adequate to accomplish this. This ratio depends on the particular receiver, but it is a smaller ratio than is required to maintain adequate reception with a mobile receiver.

Even relatively small fluctuations in signal strength due to the adding and subtracting of carriers will cause a noticeable fluctuation in receiver S/N, unless the residual signal strength in the null area is adequate to keep the receiver well into quieting.

This effect, more than any other, makes the areas of inadequate carrier ratios unlistenable. It is important to note that for the mobile receiver, the picket fencing problem is present whether or not the carriers are locked.

Modulation Effects

So far, we have dealt primarily with the effect of adding a second unmodulated carrier onto a main unmodulated carrier. At best, understanding the consequences of imperfectly correlated modulation superimposed on two carriers is difficult, if not futile. There are, however, two areas of modulation equalization which are fairly straightforward.

Time Delay Equalization

For two identical signals that have a constant time delay between them, at any (base band) frequency where the time delay is equivalent to 180° or *N* multiples of 180°, there will be complete cancellation, producing a combing effect. This is especially destructive to composite FM stereo and subcarrier performance. In fact, a 7.5 μsec group delay addition to one composite signal will cancel a 67 kHz subcarrier. Likewise, a 26.3 μsec delay will cancel the stereo pilot tone.

Unfortunately, time delay equalization cannot eliminate cancellation of modulation components in equal carrier areas, since differences in propagation times between the main and the booster will be adequate in some cases to cause the cancellation. At best, the use of group delay can move the location of the nulls relative to the transmitters. Care must be taken to insure the network exhibits a constant group delay, as nonlinear delay can seriously degrade stereo performance, especially stereo separation.

Deviation Calibration

If the frequency deviation of both modulators is not precisely calibrated, a condition of dynamic interference will occur during modulation. For example, two separate modulators are fed identical amplitude, delay equalized sine waves. The level is adjusted to produce a nominal 100% (±75 kHz) modulation. The first modulator swings the carrier exactly ±75 kHz, as predicted, but the second modulator only modulates its carrier ±74 kHz (98.67%).

Careful examination shows that the second carrier will interfere with the first in the following manner. Assuming equal carriers in an additive RF location and starting at time zero, we have two carriers of exact frequency producing a single carrier whose amplitude is derived from Equation 15. As we move positively in frequency with the modulation, the carrier frequencies diverge until, at the peak of modulation, the carriers are 1 kHz apart. From Equation 15, this produces a 1 kHz FM modulation at a B of 1 (1 kHz deviation) and an AM modulation equivalent to 100% at 1 kHz.

Actually, the detected interference is a frequency sweep from dc to 1 kHz to dc and back to 1 kHz for each complete cycle of modulation applied. Subjectively, this type of interference sounds similar to white noise, but it is only present during modulation. It is also most prevalent during the maximum modulation peaks, as this is the point of maximum carrier divergence with maximum detected loudness. This relationship of maximum interference during the periods of maximum modulation tends to mask the noise.

Correcting the Interference

Both forms of modulation related interference are most prevalent in areas of nearly equal carriers. With adequate carrier ratios, both types are effectively eliminated by the capture effect in the FM receiver. It is also important to keep the proper perspective in assessing the importance of these forms of interference. Remember, these are most prevalent in areas where, even without modulation, the carriers are adding and subtracting, tending to make the signal unlistenable as the receiver location is moved.

Booster Systems

Two main components must be present at the booster station. Station program material must be transmitted either in the form of composite stereo or as AES3 serial audio data. In addition, some form of frequency locking information must also be present.

Using a Radio Link

One method of interconnection is by the use of a radio link, such as a composite STL between the main transmitter and the booster. This method has several advantages, including high quality transmission, reliability, total signal control and economy. It is also capable of transmitting the frequency reference signal with the composite stereo via subcarrier. In this way, one radio link supplies both the station programming and reference information.

Synchronization to GPS Satellite Atomic Timebase

Some digital FM exciters include the capability to lock the FM carrier frequency and stereo pilot tone frequency to an external 10MHz reference signal derived from a Global Positioning System (GPS) satellite receiver. This high accuracy, external reference takes the place of the internal 10 MHz carrier reference in the exciter. When two or more exciters are locked to a GPS source, they become frequency-locked to a very high degree of stability. They also eliminate the need for an STL subcarrier or land line to convey a synchronizing signal since all timebases within the digital FM exciters inherit the accuracy of the atomic clock aboard the GPS satellites. In addition, the FM deviation of all the digital exciters in the network will be exactly matched if the audio is transported via an uncompressed AES3 serial data link since all digital modulators will be fed identical audio data. If the AES3 transport is through T1/E1 data lines via data format conversion, the time delay variations in the T1/E1 link must be dynamically compensated. The SynchoCast system recently introduced by Intraplex uses GPS timing to allow precision differential delay adjustments between multiple T1/E1 links and to maintain the delay compensation over time.

OPERATIONAL MEASUREMENTS

The FCC considers certain parameters to be important enough to justify almost continuous observation. Especially important are the modulation level, carrier frequency, and output power level. Consult the current revision of Part 73 of the CFR because there may have been changes since this chapter was printed.

FM Modulation Measurement

Measuring FM modulation can be accomplished with a broadcast type modulation monitor or with a so-called modulation analyzer test instrument. Some FM exciters have built-in peak modulation displays for convenient setup and adjustment. Once the initial levels are correctly set, modern audio processing equipment will usually hold the modulation levels within the desired window.

The broadcasting industry has been interested in the best method to determine the modulation percentage for complex program material. The inability of mechanical meter movements to follow short-duration, nonrepetitive peaks accurately has received special attention. Standard modulation meter movements cannot follow modulation peaks with the required accuracy. For this reason, modulation monitors have a peak-indicating device that can be preset to flash at the particular level of interest. This device should be used instead of the meter to determine peak modulation conditions of the transmitter.

The reason for setting a peak deviation limit is so the related occupied bandwidth does not increase to the point of interfering with stations on adjacent channels and causing excessive audio distortion in the FM receiver by exceeding the receiver's IF bandwidth. The FCC presently enforces the modulation limit by monitoring the instantaneous peak deviation of the station as displayed on an oscilloscope. This method of measurement does not exactly correlate with the station's occupied bandwidth because the duty-cycle of the modulation peaks is not taken into account. As a result, many sophisticated peak limiting and overshoot control devices have appeared on the market to maximize loudness without exceeding the peak deviation limit. These devices remove the low energy peaks that would extend beyond 100% modulation. The use of these devices does cause some degradation of the audio quality, they would not be used if the method of modulation measurement were changed to one based on occupied bandwidth. Recently introduced modulation measurement devices ignore short duration overshoots and provide modulation level indications that more accurately reflect the resulting occupied bandwidth. The FCC now enforces occupied bandwidth requirements for the FM band by measuring compliance with the FM emissions mask defined in Section 73.317 of the CFR.

Carrier Frequency Measurement

The average carrier frequency must be measured with an accurate frequency monitor and maintained to within ±2,000 Hz of the assigned channel. These monitors fall into two categories: analog display of the frequency error from the nominal carrier frequency and digital display of the absolute carrier frequency. The trend is toward the digital counter because of its accuracy and ease of use.

Historically, the FCC required FM stations to utilize type-approved modulation and frequency monitors, but this requirement has since been eliminated. Each station is still required to maintain its frequency, modulation and audio performance within the FCC limits defined in Part 73 of the CFR, but the responsibility for selecting the method of measurement and type of measuring equipment is now up to the station operator. Every quality conscious station will want to have the necessary equipment to accurately evaluate the signal being broadcast. New options are available in high performance yet general purpose test equipment, now that monitor type approval is no longer required. For instance, modern modulation analyzers provide frequency agility as well as greater measurement capability than the more specialized modulation monitors. General purpose frequency counters are now available with sufficient accuracy to measure the carrier, subcarrier and stereo pilot frequencies directly. Spectrum analyzers provide a wide range of capability including the measurement of harmonic and spurious frequencies at the carrier frequency, composite baseband, Bessel nulls, occupied bandwidth, stereophonic and SCA crosstalk and synchronous AM.

Measurement of RF Power Output

The methods for determining RF output power are specified in Part 73 of the CFR. An accurately calibrated directional wattmeter provides an excellent means of making a direct measurement of RF output power. Until recently, the directional wattmeter was seldom used as the primary RF power determining method because of the requirement for recalibration to a traceable standard at regular intervals. Use of the indirect method of power measurement avoided this requirement. The FCC now permits the use of the transmitter power output meter directly, if it is periodically calibrated by comparison with the indirect method, instead of with a dummy load and standard wattmeter.

Using the indirect method, the output power is calculated from a measurement of the dc input power multiplied by the efficiency factor of the final amplifier stage. The efficiency factor is provided by the transmitter manufacturer on the final test data sheet or in the instruction manual and must be applicable to the particular frequency and power level in use. The power input to the final amplifier stage is normally defined as the product of plate voltage and plate current to this stage. Multiple output stages which are combined for the total power must have their individual dc power inputs arithmetically summed to obtain the total power input.

The directional wattmeter can be used as a check when compared to the power output calculated by the indirect method to determine if the efficiency factor has changed due to incorrect tuning, changing antenna conditions or a weak output device.

Logging of Transmitter Operating Parameters

Although the FCC has deleted specific logging requirements, FM stations may still wish to log certain transmitter parameters at three-hour intervals to track equipment performance trends. The minimum suggested entries are:

- Final amplifier plate/drain voltage
- Final amplifier plate/drain current
- RF transmission line current, voltage or direct power.

The carrier frequency, stereo pilot frequency and SCA subcarrier frequency require less frequent measurement and logging. Microprocessor based control systems often provide automatic logging of all transmitter parameters, including overloads and tolerance flagging.

Proof of Performance Measurements

While the FCC no longer requires proof of performance tests, it is advisable to conduct such tests periodically to ensure the equipment is operating properly. These tests should include:

- Audio frequency response
- Audio frequency harmonic distortion
- FM signal-to-noise ratio
- AM noise level.

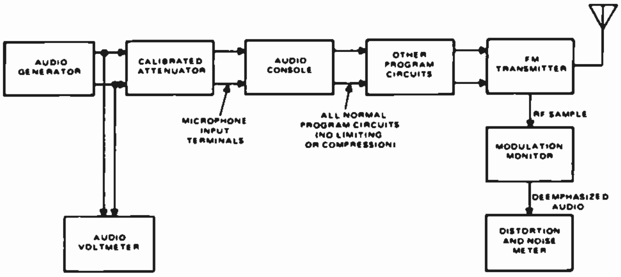


Figure 4.3-47. Block diagram of test setup for audio performance measurements.

Figure 4.3-47 is a block diagram of a typical test set-up for proof-of-performance measurements.

Measurement of Audio Frequency Response

The audio frequency response of the system is measured in reverse, that is, a constant percentage of modulation is maintained for all modulating frequencies by adjusting the amount of attenuation between the audio generator and the microphone input terminals. This is necessary because of the rising response due to preemphasis. Frequency response data is taken at three levels of modulation: 25%, 50% and 100%. The audio voltmeter which measures the audio generator output voltage is used to maintain a constant voltage level versus frequency at the generator output terminals. The precision attenuator dials are adjusted for each modulating frequency to maintain the desired modulating level, and the attenuator readings, in decibels, are recorded. Readings should be taken for the following modulating frequencies: 50, 100, 400, 1,000, 5,000, 7,500, 10,000 and 15,000 Hz. When the attenuation, in decibels, is plotted versus frequency, the 75 microsecond preemphasis curve is obtained if the system frequency response is perfect. Deviations from the ideal response are permitted to the extent that allows the measured curve to be fitted between the upper and lower limit curves shown in Figure 4.3-9(a). The procedure for doing this is to offset the measured curve by subtracting or adding the same number of decibels from each of the measured values. This process may be repeated until a fit is obtained. If it is impossible to obtain a fit within the limits by subtracting or adding the same value of attenuation from all measured values, the system frequency response is inadequate and corrections must be made.

Measurement of Audio Harmonic Distortion

Total harmonic distortion (THD) of the system from microphone input terminals to transmitter output is measured by modulating the transmitter with sinusoidal modulating signals having low distortion and observing the harmonic content at the output of the modulation monitor. For this measurement, preemphasis is used in the transmitter and deemphasis is used in the monitor. The distortion analyzer must respond to deemphasized harmonics through 30 kHz.

The type of distortion meter normally used in this test reads not only harmonic distortion but also noise in the audio passband. For this reason, THD measurements above 5 kHz may be noise limited due to the effect of deemphasis. A more accurate method of distortion measurement in the presence of noise is to use an audio frequency spectrum analyzer to determine the total RMS value of the individual distortion products.

The THD is typically measured under the following conditions:

- For modulating frequencies of 50, 100, 400, 1,000 and 5,000 Hz at modulating levels of 25, 50 and 100% modulation
- For modulating frequencies of 10 kHz and 15 kHz at a modulating level of 100%.

The preceding measurements must show that the system THD is less than 3.5% for modulating frequencies between 50 and 100 Hz, less than 2.5% for modulating frequencies between 100 and 7,500 Hz, and less than 3.0% for modulating frequencies between 7,500 and 15,000 Hz. Most modern equipment will pass these tests by a wide margin. If distortion levels greater than these are measured, the system requires adjustment or repair. Tuning the transmitter for minimum even order harmonic distortion will result in a symmetrical group delay response and optimum FM modulation performance. This can be accomplished either by observing the even order harmonics in the demodulated baseband or by placing an audio bandpass filter tuned to the second harmonic or other even order harmonic of the modulating frequency ahead of the audio distortion analyzer. If the transmitter is modulated by a single monaural tone of 9.5 kHz, the 19 kHz (2 × 9.5 kHz) bandpass filter in the stereo modulation monitor's pilot injection metering circuit can be used to tune the transmitter for minimum second harmonic distortion.

Measurement of Two-Tone Intermodulation Distortion

Audio intermodulation distortion (IMD) measurements are a quick and accurate way to evaluate the system performance before a complete set of THD measurements are made. If the system will pass a single 60 Hz/7 kHz, 4:1 (Society of Motion Picture and Television Engineering (SMPTE)-IMD) measurement at 100% modulation, it will probably pass all of the THD measurements up to 10 kHz. This dual frequency test signal (SMPTE-IMD) is on track 30 of the *NAB Broadcast and Audio System Test CD.*

Other types of difference-tone, swept two-tone and sinewave/squarewave IMD measurements can reveal more subtle problems in the system such as transient IMD due to insufficient audio feed back bandwidth in the audio amplifier stages.

Measurement of FM S/N

The FM signal-to-noise ratio of the system is also measured from the microphone input terminals to the transmitter output. The residual noise level at the monitor output is measured with an audio voltmeter. For this measurement, preemphasis is employed in the transmitter and deemphasis in the monitor. The residual audio noise level is referenced to the signal level produced by 400 Hz (L + R, mono) modulation at the 100% level (±75 kHz deviation).

The procedure for making the FM S/N measurement is:

- Modulate the transmitter with a 400 Hz sine wave applied at the microphone input terminals of the console and set the level for 100% modulation
- Read and record the audio signal level appearing at the modulation monitor output terminals. If the monitor has audio metering capability, the meter gain should be set for a 0.0 dB reference level according to the manufacturer's instructions

- Remove the modulation and terminate the console audio input terminals with a resistor equal to the normal microphone output impedance
- Read and record the residual audio noise voltage in decibels below the 400 Hz reference signal level. The measured S/N must be at least 60 dB. State-of-the-art equipment can provide a S/N of better than 90 dB, so it is relatively easy to meet the FCC requirement of 60 dB.

Measurement of AM S/N

The perfect FM transmitter will have an absolutely constant output, regardless of FM modulation or power supply variations. In practice, there is always some residual amplitude modulation of the FM transmitter. Two types of AM S/N are of interest to the FM broadcast engineer:

1. Asynchronous AM S/N measured without FM modulation is required by the FCC Rules and is primarily related to power supply ripple.
2. Synchronous AM S/N or incidental carrier AM (ICAM) measured with FM modulation is not required by the FCC Rules and is related to the tuning and overall bandwidth of the system.

Asynchronous AM

Residual amplitude modulation of the transmitter output without FM modulation, due primarily to power supply ripple, is measured with an AM envelope detector. Most FM modulation monitors include an AM detector for this purpose. The detector must include 75 microsecond deemphasis of its output. AM noise measurements must be made directly at the transmitter output (or an accurate sample of its output). No amplifying or limiting equipment may be used between the transmitter output and the AM detector since this equipment would modify the residual AM noise level present. The FCC Rules require residual AM noise to be 50 dB below the level which would represent 100% amplitude modulation of the carrier. Since the transmitter cannot be amplitude modulated, this reference must be established indirectly by a measurement of the RF carrier voltage. (Refer to the instructions of the detector manufacturer to determine the reference level. Generally, the reference level is determined by setting a carrier level meter to a specified reading. If the transmitter is unable to meet the 50 dB requirement, the problem can usually be traced to a power supply component or to line imbalance in a three phase system.)

Synchronous AM

Synchronous AM is a measure of the amount of incidental amplitude modulation introduced onto the carrier by the presence of FM modulation. Although this measurement is not required by the FCC Rules, it provides information about the amplitude response and tuning of the transmitter. Measurement of synchronous AM also gives the station engineer an idea of the overall system bandwidth and whether the passband is positioned correctly. Since all transmitters have limited bandwidth, there will be a slight drop-off in power output as the carrier frequency is swept to either side of the center frequency. This slight change in RF output level follows the waveform of the signal being applied to the FM modulator causing AM modulation in synchronization with the FM modulation. The concept is similar to the slope detection of FM by an AM detector used in conjunction with a tuned circuit.

Synchronous AM measurements are made directly at the transmitter output (or an accurate sample of its output). No amplifying or limiting equipment may be used between the transmitter output and the AM detector since nonlinearities in this equipment could modify the synchronous AM level present. Since the transmitter cannot be fully amplitude modulated, an equivalent reference level must be established indirectly by a measurement of the RF carrier voltage. (Refer to the instructions of the detector manufacturer to determine this reference level. Generally, the reference level is determined by setting a carrier level meter to a specified reading or to obtain a specific dc voltage level at the output of the detector diode without modulation. Care must be taken when making these measurements that the test setup does not introduce synchronous AM and give erroneous readings which would cause the operator to mistune the transmitter to compensate for errors in the measuring equipment.)

The input impedance of the envelope detector must provide a nearly perfect match so that there is a very low VSWR on the sampling line. Any significant VSWR on the sampling line will produce synchronous AM at the detector because the position of the voltage peak caused by the standing wave moves along this line with FM modulation. A thru-line type of directional coupler normally used to drive the wattmeter movement has the envelope detector diode built into the sampling element. This provides a dc component that the meter movement responds to plus the demodulated synchronous AM component that the meter movement does not respond to. If the thru-line element output is fed to an oscilloscope instead of the wattmeter movement, the synchronous AM waveform can be accurately measured. This approach eliminates the errors due to VSWR on the sampling line, since the detector is located at the sampling point. The manufacturer of the thru-line coupler can supply the special connectors and/or cables to connect its output to the oscilloscope. Care must be taken to avoid hum pick-up from ac ground loops while making these low level measurements. Both the thru-line element detector and the precision envelope detectors have some residual RF on their dc output, so an RF filter network may be required between the detector and the input the oscilloscope.

Most FM demodulators cannot be relied upon to make accurate synchronous AM noise measurements, so it is a good idea to cross-check the demodulator reading directly against the demodulated output of a precision envelope detector. This can be done by first measuring the dc component of the waveform with a voltmeter or by dc coupling the scope input. The scope is then ac coupled, and the input sensitivity is increased until an accurate peak to peak measurement of the ac modulation component can be made. The peak-to-peak ac voltage is then divided by twice the dc component to obtain the voltage ratio. Twenty times the log (base 10) of the voltage ratio is the actual synchronous AM noise level in dB below equivalent 100% AM modulation. Multiplying the voltage ratio by one hundred yields the percent of AM modulation. Note that the peak detected value of the carrier must be doubled to convert it to the peak-to-peak value of the carrier. The ratio of the peak-to-peak modulation component to the peak-to-peak carrier is then used to calculate the percentage of synchronous AM modulation.

What is an Acceptable Level of Synchronous AM?

Synchronous AM of 35 dB or more below equivalent 100% AM is considered to be acceptable since the limited bandwidth of the IF filter in the receiver will reintroduce higher levels of synchronous AM to the FM signal before demodulation. Higher levels of synchronous AM can cause increased "chopping" of the signal at the receiver near limiting threshold under weak signal fringe area conditions and can exacerbate multipath problems. Excessive synchronous AM is also an indirect indication of passband-induced distortion problems that degrade stereo performance and SCA crosstalk.

Many older multitube transmitter designs presently in use will have as much as 6% (−30 dB) synchronous AM when simply tuned for best power output and efficiency even though the asynchronous AM (without modulation) may be better than −50 dB. Some of the newer single tube transmitters can be adjusted for 50 dB or more suppression of synchronous AM. The synchronous AM level of virtually any FM transmitter can be improved by proper tuning techniques. An approximation to the overall system bandwidth can be related to the synchronous AM as shown in Table 4.3-3.

Limitations of Synchronous AM Measurements

Synchronous AM measurements are an indirect way of evaluating and optimizing FM performance. Even though synchronous AM measurements are a helpful aid to correctly tune an FM transmitter, these measurements tell only the amplitude response half of the total story. Transmitter tuning also affects the group delay (time) response which in turn affects the relative time delays of the higher order FM sidebands. Even though the amplitude response appears flattened when the grid is heavily driven, the group delay (time) response still has a serious effect on the higher order FM sidebands.

Table 4.3-3

Approximate system bandwidth as related to synchronous AM.

Synchronous AM (Below 100% AM) (+75 kHz@1 kHz FM)	Approximate Bandwidth of Transmitters (−3 dB)	RF Level Variation at Receiver Limiter (%)	(dB)
−30 dB	410 kHz	6.32%	0.57 dB
−35 dB	550 kHz	3.54%	0.31 dB
−40 dB	730 kHz	2.00%	0.18 dB
−45 dB	1.00 MHz	1.12%	0.10 dB
−50 dB	1.34 MHz	0.64%	0.06 dB
−55 dB	1.82 MHz	0.36%	0.03 dB
−60 dB	2.46 MHz	0.20%	0.02 dB

Synchronous AM versus Symmetrical Group Delay Response

Computer simulations as well as empirical measurements made on FM transmitters showed that group delay asymmetry results in much more distortion than asymmetrical amplitude response. [23] As long as the group delay response is symmetrical, the amount of synchronous AM will have little effect on the FM modulation performance and distortion. Most FM transmitters will exhibit a significant increase in synchronous AM when tuned for symmetrical group delay response even though this results in best FM modulation performance. Tuning for minimum synchronous AM is a good starting point, but it is more desirable to finish tuning at the symmetrical group delay point.

Fine tuning the input and output for minimum even-order harmonic distortion will optimize the group delay (time) response. Transmitters that utilize wideband solid-state IPAs will add less distortion to the FM signal because both the amplitude and group delay (time) response will be better than systems utilizing several tuned stages.

Tuning The Transmitter for Best Performance

If properly adjusted, modern power amplifiers discussed in the preceding sections can operate with high reliability and power efficiency without compromising subcarrier performance. All optimization should be done with the APC system disabled so that the APC will not chase the adjustment to keep the output power constant. The transmitter should be connected to the normal antenna system rather than to a dummy load. This is because the resistance and reactance of the antenna will be different from the dummy load, and the optimum tuning point of the transmitter will shift between the two different loads. The tuning sequence is described below.

Initial Tuning and Loading

The transmitter is first tuned for normal output power and proper efficiency according to the manufacturer's instructions. The meter readings should closely agree with those listed on the manufacturer's final test data sheet if the transmitter is being operated at the same frequency and power level into an acceptable load.

Input Tuning and Matching

The input tuning control should first be adjusted for maximum grid current and then fine-tuned interactively with the input matching control for minimum reflected power to the driver stage. Note that the point of maximum grid current may not coincide with the minimum reflected power to a solid-state driver. This is because a solid-state driver may actually output more power at certain complex load impedances than into a 50 Ω resistive load. The main objective during input tuning is to obtain adequate grid current while providing a good match (minimum reflected power) to the coaxial transmission line from the driver. In the case of an older transmitter with a tube driver integrated into the grid circuit of the final amplifier, the driver plate tuning and the final grid tuning will be combined into one control that is adjusted for maximum grid current.

Output Tuning

The output tuning control adjusts the resonant frequency of the output circuit to match the carrier frequency. As resonance is reached, the plate current will drop while both the output power and screen current rise together. Under heavily loaded conditions this dip in plate current is not very pronounced, so tuning for a peak in screen current is often a more sensitive indicator of resonance.

Output Loading

There is a delicate balance between screen voltage and output loading for amplifiers utilizing a tetrode tube. Generally, there is one combination of screen voltage and output loading where peak efficiency occurs. At a given screen voltage, increasing the amplifier loading will result in a decrease in screen current, while a decrease in loading will result in an increase in screen current. As the screen voltage is increased to get more output power, the loading must also be increased to prevent the screen current from reaching excessive levels. Further increases in screen voltage without increased loading will result in a screen overload without an increase in output power.

Automatic Power Control Headroom

APC feed back systems are utilized in many transmitters to regulate the power output around a predetermined setpoint with variations in ac line voltage or changes in other operating parameters. Modern FM broadcast transmitters may utilize a high gain tetrode as the final amplifier stage with adjustment of the screen voltage providing fine adjustment of the output power.

For each power output level there is one unique combination of screen voltage and output loading that will provide peak operating efficiency. If the screen voltage is raised above this point without a corresponding increase in loading, there will be no further increase in power output with rising screen voltage and screen current. If the screen voltage is raised without sufficient loading, a screen current overload will occur before the upward adjustment in power output is obtained. To avoid this problem, tune the transmitter with slightly heavier loading than necessary to achieve the desired power output level and allow for about 5% headroom in adjustment range. The output loading can be adjusted for a peak in output power of 5% over the desired level, and then the screen voltage can be reduced enough to return to the desired level. This procedure will allow headroom for an APC system controlling screen voltage and will result in about a 1% compromise in efficiency, but it will ensure the ability to increase power output up to 5% without encountering a screen overload.

Centering the Passband

A simple method for centering the transmitter passband on the carrier frequency involves adjustment for minimum synchronous AM. If the bandpass is narrow or skewed, increasing synchronous amplitude modulation of the carrier will result. A typical adjustment procedure is to FM modulate 100% at 1 kHz and fine-adjust the transmitter's grid tuning and output tuning controls for minimum 1 kHz AM modulation as detected by a wideband envelope detector (diode and line probe). One kHz is used as the FM modulating frequency rather than 400 Hz so that the audio highpass filter in the audio analyzer can be used to eliminate the ac line frequency related asynchronous component from the synchronous AM component. It is helpful to display the demodulated output from the AM detector on an oscilloscope while making this adjustment. Note that as the minimum point of synchronous AM is reached, the demodulated output from the AM detector will double in frequency to 2 kHz. This is because the fall-off in output power is symmetrical about the center frequency causing the amplitude variations to go through two complete cycles for every one FM sweep cycle (see Figure 4.3-48). It should be possible to minimize synchronous AM while maintaining output power and efficiency in a properly designed power amplifier.

Effect of Transmitter Tuning on the FM Side Bands

The higher-order FM side bands will be slightly attenuated in amplitude and shifted in time (group delay) as they pass through the final amplifier stage. The alterations in the side band structure that are introduced by the amplifier passband result in distortion after FM demodulation at the receiver. The amount of

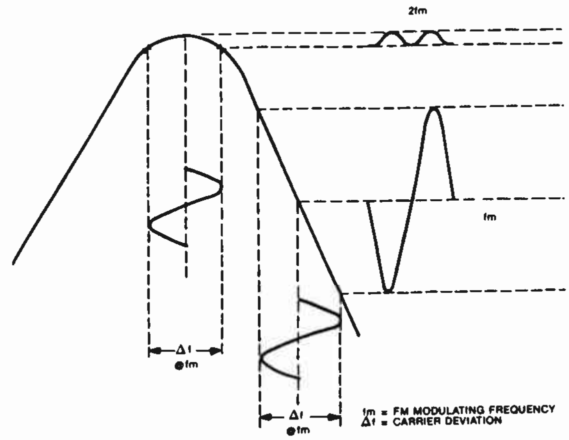


Figure 4.3-48. Synchronous AM waveforms.

distortion depends on the available bandwidth versus the modulation index being transmitted. For a given bandwidth limitation, the distortion can be minimized by centering the passband of the amplifier around the signal being transmitted. This will cause the amplitude and group delay errors to affect both the upper and lower sidebands equally (symmetrically). Tuning an amplifier for minimum plate current or for best efficiency does not necessarily result in a centered passband. One way to center the amplitude passband is to tune the amplifier for minimum synchronous AM modulation while applying FM modulation to the transmitter. Since the circuit topology of most transmitters exhibits a difference in tuning between the symmetrical amplitude response and the symmetrical group delay response, FM modulation performance can be further improved by tuning for symmetrical group delay rather than for minimum synchronous AM. The symmetrical group delay tuning point usually does not coincide exactly with the symmetrical amplitude tuning point but rather falls between the point of minimum synchronous AM and the point of maximum efficiency.

The transmitter may be tuned for minimum intermodulation distortion in left-only or right-only stereo transmissions. Stereo separation will also vary with tuning. For stations employing a 67 kHz SCA, transmitter tuning becomes very critical to minimizing crosstalk into the SCA. Modulate one channel only on the stereo generator to 100% with a 4.5 kHz tone. This will place the lower second harmonic (L − R) stereo sideband on top of 67 kHz SCA. Activate the SCA at normal injection level without modulation on the SCA. Tune the transmitter for minimum output from the SCA demodulator. This adjustment can also be made by listening to the residual SCA audio while normal stereo programming is being broadcast.

A more sensitive test is to tune for minimum even order harmonic distortion which will result in a symmetrical group delay response and will optimize distortion, separation and crosstalk.

Modern power amplifiers have been designed to operate without compromising subcarrier performance. By providing broad band matching circuits, adjustment of these transmitters for optimum FM modulation performance (minimum distortion, minimum crosstalk, maximum separation etc.) is very repeatable and stable.

The following field adjustment techniques are listed in ascending order of sensitivity:

- Tune for minimum synchronous AM noise
- Tune for minimum IMD in the left or right channel only
- Tune for minimum crosstalk into the unmodulated SCA subcarrier
- Tune for minimum even order harmonic distortion (symmetrical group delay). (See page 000)

In any of these tests, the grid tuning is frequently more critical than the plate tuning. This is because the impedance match into the input capacitance of the grid becomes the bandwidth limiting factor. Even though the amplitude response appears flattened when the grid is heavily driven, the group delay (time) response has a serious effect on the higher order FM sidebands. For more information about tuning an FM transmitter for optimum performance, see [14, 15, 20, and 22].

OPTIMUM TUNING VERSUS EFFICIENCY

VHF amplifiers often exhibit a somewhat unusual characteristic when tuning for maximum efficiency. The highest efficiency operating point does not exactly coincide with the lowest plate current because the power output continues to rise on the inductive side of resonance coming out of the dip in plate current. If the amplifier is tuned exactly to resonance, the plate load impedance will be purely resistive, and the load line will be linear. As the output circuit is tuned to the inductive side of resonance, the plate load impedance becomes complex, and the load line becomes elliptic instead of linear since the plate current and plate voltage are no longer in phase. Apparently, best efficiency occurs when the phase of the instantaneous plate voltage slightly leads the plate current. This effect is believed to be caused by the nonlinear gain characteristics of the power amplifier tube operating on an elliptic load line.

INSTALLATION CONSIDERATIONS

Adequate planning and care in the installation of an FM broadcast transmitter and associated equipment will help avoid many problems that may be difficult and expensive to correct later. For example, poor grounds and ground loops may cause high noise levels.

Wiring the Transmitter Plant

Separate metallic shielded conduits or troughs should be provided for the audio and the ac wiring. A third conduit should be used if computer logic levels are employed for equipment control. These conduits or wiring troughs may be either overhead or below the cabinets. The ac wiring should be well separated from the audio pairs to prevent the induction of unwanted hum and noise into the audio circuits.

Audio shields should be grounded at only one point to prevent ground loops in the shields. This point may have to be found experimentally to give the lowest noise pickup. The equipment racks and transmitter should be connected together by copper straps at least two in. wide, tied to a good earth ground at one point. If a good ground screen is not available, a satisfactory ground can be provided by driving four or five copper ground rods 8-10 ft long into the ground with a spacing of about three ft. These ground rods should be tied together with a wide copper strap. The straps connecting the equipment to the earth ground should be as short and direct as practical.

It is often difficult to remove VHF-RF from the equipment by grounding because at FM carrier frequencies nearly any connection to an earth ground has an appreciable impedance. The best way to keep RF out of sensitive low level circuits is by keeping them enclosed within an RF shield and by filtering leads that enter the shielded unit when necessary. Filters in the audio lines may be made up of small bi-filar RF chokes and disc capacitors.

For stereo transmission, it is necessary to keep the L and R audio lines phased properly. To insure proper monaural compatibility, correct audio polarity must be maintained throughout the station from the microphones, tape machines and turntables through all of the audio equipment to the stereo generator audio input terminals. Stereo phone line pairs or separate RF studio-to-transmitter links (STL) should also be checked for proper polarity and equal phase delay.

The transmitter equipment should be located and arranged to provide sufficient clearance around the front, sides and rear for easy access during servicing and maintenance.

Transmitter Cooling

Almost all FM broadcast transmitters require forced air cooling to remove heat from the output stage and other assemblies within the cabinet. A very important consideration in locating the transmitter is the provision for adequate cooling air. Assuming that the overall efficiency of the transmitter is about 50%, the transmitter will generate about the same number of kilowatts of heat as it does RF power output.

Figure 4.3-49(a) shows a transmitter located in an air conditioned room. This type of closed-loop system requires no special ducting and has the advantage that the transmitter intake air is usually much cleaner than outside air. The transmitter exhaust air places a substantial heat load on the air conditioner during the summer, but it becomes a source of heat in the winter. The transmitter manufacturer can usually supply data on the number of cooling BTUs required, so that the proper size air conditioner can be selected. This method is frequently used with the lower power transmitters. A protective system should be provided to prevent over heating of the transmitter if the air conditioner fails.

Figure 4.3-49(b) shows a transmitter located in a wall separating an air conditioned room and a ventilated, but not air conditioned room. A large exhaust fan is provided in the ceiling to remove the rising hot air while an adequate cool air intake is provided in the lower portion of an outside wall. Adequate air filtering is required to keep the transmitter interior clean.

Figure 4.3-49(c) shows a transmitter located in an air conditioned room with intake and exhaust air ducts to the outside. An auxiliary blower or fan is normally required to overcome pressure drop in the ducting. This type of system requires careful design to make sure the air flow through the transmitter is not impeded by the duct work. Additional air interlocks may be required to protect the transmitter from a failure of the external fan. The air intake and exhaust openings to the outside should be provided with rain shields, insect screens and dust filters as dictated by the environment.

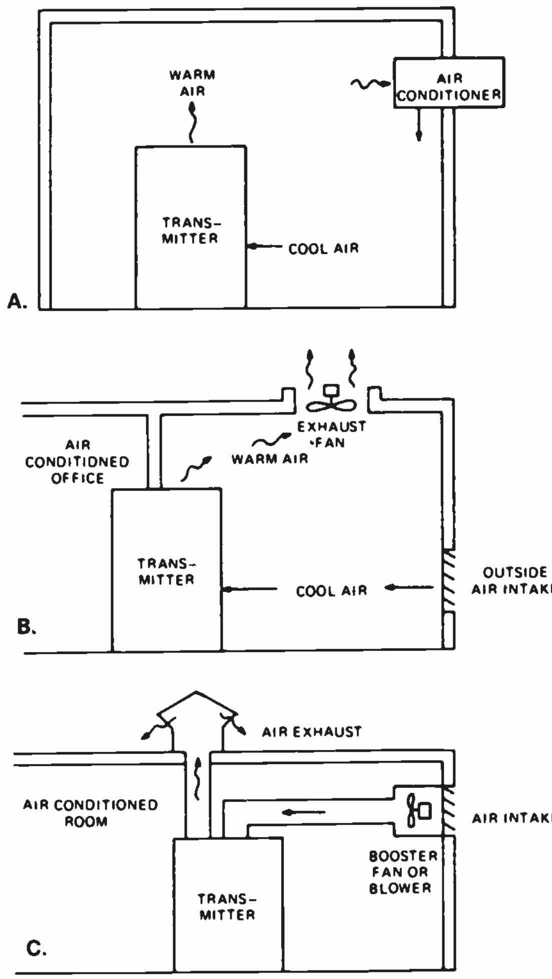


Figure 4.3-49(a–c). Three methods of providing cooling air for the transmitter.

The location of the air intake and exhaust openings should be arranged so that wind pressure will not impede the air flow.

Air filters should be periodically cleaned or replaced according to the transmitter manufacturer's instructions. This is very important because dust- or insect-clogged air filters may reduce the cooling air flow enough to cause overheating of some components. The probability of component failure increases very rapidly when cooling is insufficient. Particular attention should be paid to removing dirt and dust from high voltage components during regular maintenance after all power is removed and all components are discharged.

Dust should be cleaned from the transmitter with a suitable brush and vacuum cleaner or as otherwise recommended by the transmitter manufacturer. Usually weekly cleaning is sufficient.

Care of Power Tubes

The operating life of high power vacuum tubes can be extended by proper care. Most high power tubes utilize a directly heated cathode composed of a thoriated tungsten filament structure.

The key points to extending the life of RF power tubes are:

- Store tubes upright (upside-down or right-side-up) along the axis of symmetry, not on their side, to keep the internal elements concentrically aligned
- Use care when handling tubes to prevent mechanical shocks to the delicate internal structure. Don't set a tube on a hard surface without padding
- Keep the tube seals and anode cooler free of dust and dirt by weekly cleaning even in a clean environment
- Keep a spare tube on hand and rotate the tubes every few months to help keep the chemical "gas-getter" active so that the tubes remain gas free
- Keep a regular record of all tube operating parameters so that any trend of changes will be noticeable. If a tube fails during the warranty period, this data will be necessary to receive credit on a replacement tube
- Monitor the filament voltage on a true RMS responding instrument and log any changes for future reference. The sampling point for this voltage measurement should be located as close to the tube's filament contacts as possible to minimize errors due to voltage drops in the filament wiring.

Even a tube that has been properly operated should gradually lose emission from the cathode until it is no longer useful because the emissive material is gradually consumed. The carcass of the tube can then be rebuilt with a new cathode and recycled back into service. Tube life is not directly related to plate dissipation (within the ratings) but is related to the filament operating temperature (filament voltage) and the current density (milliamperes per heater watt) emitted by a given size filament. This means that operating a given tube type at a lower filament voltage and plate current will proportionately increase the life of the tube. For directly heated thoriated tungsten filaments, the plate current should be less than 4 milliamperes per watt of heater power for extended life.

Normally a new tube will deliver full output at a reduced filament voltage. By operating the tube at the optimum filament voltage, the filament life can be significantly extended. The optimum value may be found by slowly reducing the filament voltage from the manufacturer's rated value until the RF power output drops about 2% and then increasing the filament voltage until the RF power increases back up 1%. Recent informational bulletins on extending tube life from tube manufacturers permit the filament voltage to be reduced more than 5% below the manufacturer's rating as long as operation is closely monitored to stay above the point where there is a 1% drop in power output. [26] A brand new tube should be operated at the full rated filament voltage for the first 200 hours before the voltage is reduced to the optimum value for long life. This will assure that the gas-getter is properly activated. As the tube ages, the filament voltage will have to be increased to stay at the optimum value, until RF output power cannot be maintained at or above the rated value of filament voltage. At this point, the tube's useful life comes to an end. Check the manufacturer's data sheets and most recent application notes which are often enclosed with the tube for detailed information. An excellent guide to the proper care of power tubes is listed in references [10 and 26]. [17] gives detailed information about how to specify a proper forced-air cooling system for power tubes.

Preventive Maintenance

Preventive maintenance is equipment inspection and maintenance performed at regular intervals before an operational problem develops. The long term benefits are great because potential problems are discovered and solved while they are still easily manageable. A checklist of a few typical preventive maintenance items for an FM transmitter plant might include:

- Weekly overall internal and external cleaning and inspecting for damage or excessive wear
- Lubricate motors, tuning gears, and other moving parts at intervals recommended by the manufacturer
- Checking and logging all meter readings, including filament voltage, daily. Then compare these readings with the previous set of readings as an aid to diagnosing a developing problem
- Regularly exercising the automatic power control and any other servo systems.
- Checking the antenna lighting and de-icer systems
- Checking the transmission line pressurization and VSWR
- Checking all air filters in the transmitter plant and clean or replace as required
- Checking the proper operation of all monitoring and remote control equipment.

In addition, good overall housekeeping will pay big dividends in the long run by keeping equipment clean and free of problems that would otherwise be caused by dirt build-up.

Maintenance Systems

The key to making any maintenance program work is to set up formal checklist, logging, parts inventory management and repair scheduling systems. These systems provide the discipline required to keep the maintenance routine accurate and complete. Each station should develop a system suited for the particular physical plant involved. When there is more maintenance and repair work needing attention than there is time to do it all, set priorities for completing each item so no item gets forgotten. Accurate notebooks describing all installation and maintenance work are a very helpful part of any maintenance system especially when work spans many years.

REFERENCES

1. Lessing, Lawrence, *Man of High Fidelity: Edwin Howard Armstrong*, Bantam Publishing, The Armstrong Foundation, Columbia University, New York, NY, 1959.

2. *Reference Data for Radio Engineers, Fifth Edition,* ITT-Howard W. Sams and Co., Inc., New York, NY, 1970.
3. Terman, Frederick E., *Electronic and Radio Engineering,* McGraw-Hill Book Company, New York, NY, Fourth Edition, 1955.
4. Clarke, Kenneth K. and Hess, Donald T., *Communications Circuits: Analysis and Design,* Addison-Wesley Publishing Company, Reading, MA, Second Printing, 1978.
5. Rohde, Ulrich L., *Digital PLL Synthesizers-Theory and Design,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1983.
6. Krauss, Herbert L., Bostian, Charles W., and Raab, Frederick H., *Solid-State Radio Engineering,* John Wiley and Sons, New York, NY, 1980.
7. Uyttendaele, Antoon G., *Design Requirements and Operational Features of the Gates Dualtran RF Switching System,* Harris Corporation - Broadcast Division, Quincy, IL.
8. Johnson, Joseph, *Solid Circuits,* Acrian/Communications Transistor Company, Cupertino, CA, 1973.
9. Howe, Harlan Jr., *Simplified Design of High Power, N-Way, In-Phase Power Divider/Combiners,* Microwave Journal, December, 1979.
10. *Care and Feeding of Power Grid Tubes,* Eimac Division of Varian Corporation, San Carlos, CA, 1967.
11. Hershberger, David and Weirather, Robert, *Amplitude Bandwidth, Phase Bandwidth, Incidental AM, and Saturation Characteristics of Power Tube Cavity Amplifiers for FM,* Harris Corporation Broadcast Division, Quincy, IL, 1982.
12. *Motorola Silicon Rectifier Manual,* Semiconductor Group-Motorola Inc., Phoenix, AZ, 1980.
13. Hnatek, Eugene R., *Design of Solid-State Power Supplies,* Van Nostrand Reinhold Company, New York, NY, Second Edition, 1981.
14. Mendenhall, Geoffrey N., *A Study of RF Intermodulation Between FM Broadcast Transmitters Sharing Filterplexed or Co-located Antenna Systems,* Broadcast Electronics Inc., Quincy, IL, (c) 1983.
15. Mendenhall, Geoffrey N., *The Composite Signal-Key to Quality FM Broadcasting,* Broadcast Electronic Inc., Quincy, IL, (c) 1981.
16. Lyles, John T.M. and Shrestha, Mukunda B., *Transmitter Performance Requirements for Subcarrier Operation,* Broadcast Electronics Inc., Quincy, IL, (c) 1984.
17. Steinkamp, Jeffrey H., *Air Flow Measurements Using a Variable Frequency AC Generator,* Broadcast Electronics Inc., Quincy, IL, (c) 1984.
18. Robinson, John T. and Whipple, L. Glenn, "FCC Regulations Governing the Audio Fidelity Characteristics of FM Broadcast Equipment," *Audio Engineering Society Journal,* Vol. 16, No. 3, July 1968.
19. Bruene, Warren B., "An Inside Picture of Directional Wattmeters," American Radio Relay League—*QST* magazine, April 1959.
20. Mendenhall, Geoffrey N., *Techniques for Measuring Synchronous AM Noise in FM Transmitters,* Broadcast Electronics, Inc., Quincy IL, (c) 1988.
21. Anthony, Edward J., *Practical Considerations for the Implementation of a Reliable Synchronous FM Booster,* Broadcast Electronics, Inc., Quincy, IL, (c) 1988.
22. Shrestha, Mukunda B., *The Significance of RF Power Amplifier Circuit Topology on FM Modulation Performance,* Broadcast Electronics, Inc., Quincy, IL, (c) 1990.
23. "FMSIM," *FM Stereo Simulation and Analysis Program by Quantics,* 10373 Pine Flat Way, P.O. Box 2163, Nevada City, CA 95959-9136, (c) 1990.
24. Anthony, Edward J., *Optimum Bandwidth for FM Transmission,* Broadcast Electronics, Inc., Quincy, IL, (c) 1989.
25. Shrestha, Mukunda B., *Personal Safety Considerations with Broadcast Transmitters,* Broadcast Electronics, Inc., Quincy, IL, (c) 1989.
26. Artigo, Robert, "Extending Transmitter Tube Life," *Broadcast Management/Engineering Magazine,* March 1982, Revised March 1990 and Reprinted as Eimac Application Bulletin AB-18, Eimac Division of Varian Corporation, San Carlos, CA.
27. Twitchell, Edwin R., *A Digital Approach to an FM Exciter,* IEEE Transactions on Broadcasting, 1991.
28. AES3 Technical Standard Audio Engineering Society, Inc. *AES Recommended Practice For Digital Audio Engineering: Serial Transmission Format For Two-Channel Linearly Represented Digital Audio Data.* For more information, refer to: AES3-199
29. Bytheway, David L., "Charting a Path Through the Maze of Digital Audio Technology," *Broadcast Engineering Magazine,* July 1991.
30. Fry, Richard J., *Harris DIGIT FM Exciter Facts and Features,* Harris Corporation-Broadcast Division, Quincy, IL, (c) 1995.
31. Mendenhall, Geoffrey N., *Implementing an Uncompressed Digital Path from the Studio to the On Air Signal,* Harris Corporation-Broadcast Division, Quincy, IL, (c) 1997.
32. Dittmer, Tim W., *Advances in Digitally Modulated RF Systems,* Harris Corporation-Broadcast Division, Quincy, IL, (c) 1997.

4.4
FM STEREO AND SCA SYSTEMS

JOHN KEAN
MOFFET, LARSON & JOHNSON, INC., FALLS CHURCH, VA

INTRODUCTION

Stereophonic sound FM broadcasting and Subsidiary Communications Authorization (SCA) services are a form of multiplexing that had its origins over 60 years ago, when high-fidelity audio and facsimile messages were simultaneously transmitted from the Empire State Building in New York City to experimental receive sites in New Jersey.[1] Those historic multiplexing efforts demonstrated the value of wideband frequency *modulation*, allowing modern FM broadcasting to provide services to the public of both quality and variety.

Despite its illustrious start, FM broadcasting in the 1950's was a marginal business. The number of operating FM's declined annually from about 700 in 1947 to slightly over 500 in 1957. Many depended on additional income from their SCA services, which offered background music and other programs to stores and offices over medium fidelity audio subcarriers. The FCC gave the then struggling industry a boost by granting something AM and TV didn't have—stereo. The FCC's April 1961 ***Report and Order*** authorized transmission of a stereophonic sound system that combined the system proposals of Zenith Radio Corp., Chicago and General Electric Co.

FM broadcasters did not show much excitement for their new capability, but the growing availability of stereophonic LP records and home stereo equipment created a natural market for FM stereo. Most agree that FM's resurgence and eventual dominance as an audio medium was due in large part to stereophonic transmission. FM licensees, as well, still actively employ SCA.

THE COMPOSITE BASEBAND

This section contains definitions of the terms commonly used in the FM stereo and SCA systems.[2] Since some of these terms are misused or ambiguous, the intention here is to establish meanings that will be used throughout this chapter.[3] The list was selected to clarify certain terms and is not a complete glossary.

Key Terms For Stereo and SCA

Multiplexing. In its simplest sense, multiplexing implies that two or more independent sources of information are combined for carriage over a single medium, namely, the radio frequency *carrier*, and then are separated at the receiving end. In stereophonic broadcasting, for example, program information consisting of left and right audio signals are multiplexed onto an FM carrier for transmission to receivers which subsequently recover the original audio signals.

Channel. A transmission path. The usage herein distinguishes between the concept of a channel and a signal (main channel, stereophonic subchannel, etc., and left and right audio *signals*).

Composite Baseband Signal. A signal which is the sum of all signals that frequency modulate the main carrier. The signal includes the main channel signal, the modulated stereophonic subcarrier, the pilot subcarrier and the SCA subcarriers.

FM Baseband. The frequency band from 0 Hz to a specified upper frequency which contains the composite baseband signal.

Main Channel. The band of frequencies from 50 (or less) Hz to 15,000 Hz on the FM baseband which contains the main channel signal.

Main Channel Signal. A specified combination of the monophonic or left and right audio signals which frequency modulates the main carrier.

Stereophonic Sound. The audio information carried by a plurality of channels arranged to afford the listener as sense of the spatial distribution of sound sources. Stereophonic sound includes, but is not limited to, biphonic (two channel), triphonic (three channel) and quadraphonic (four channel) services.

Stereophonic Sound Subchannel. The band of frequencies from 23 kHz to 99 kHz (53 kHz for two channel transmission) containing sound subcarriers and their associated sidebands.

Subchannel. A transmission path specified by a subchannel signal occupying a specified band of frequencies.

Subchannel Signal. Subcarriers and associated sidebands which frequency modulate the main carrier. It is synonymous with subcarrier, as in the stereophonic subcarrier or SCA subcarrier.

Frequency Deviation. The peak difference between the instantaneous frequency of the modulated wave and the average carrier frequency.

Percentage Modulation. The ratio of the actual frequency swing of the carrier to the frequency swing defined as 100% modulation, expressed in percentage. Although current FCC rules conditionally permit greater than 100% modulation when SCAs are transmitted, a frequency swing of ±75 kHz is still defined as 100 percent modulation.

Injection. The ratio of the frequency swing of the FM carrier by a subchannel signal to the frequency swing defined as 100% modulation, expressed in percentage. The total injection of more than one subchannel signal is the arithmetic sum of each subchannel injection.

Crosstalk. An undesired signal occurring in one channel caused by an electrical signal in another channel.

Linear Crosstalk. A form of crosstalk in which the undesired signals are created by phase or gain inequalities in another channel or channels. Such crosstalk may be due to causes external to the stereophonic generator; consequently it is sometimes referred to as system crosstalk.

Nonlinear Crosstalk. A form of crosstalk in which the undesired signals are created by harmonic distortion or intermodulation of electrical signals in another channel or channels. Such crosstalk may be due to distortion within the stereophonic generator or FM transmitter; consequently it is sometimes referred to as transmitter crosstalk.

Frequency Spectrum And Modulation Limits

The FCC's stereophonic transmission standards are contained in Section 73.322 of the FCC Rules, and the SCA transmission standards are contained in Section 73.319. Readers should refer to the Commission's rules for a listing of its standards.

The composite baseband extends to 99 kHz and may be used in support of either stereophonic or SCA multiplex services. Stereophonic broadcasting includes, but is not limited to, biphonic service. Within the frequency range of 23 kHz to 99 kHz any form of amplitude modulation (DSB, SSB, etc.), angle modulation (FM or PM) or frequency shift keying of a multiplex subchannel is permitted.

An actual authorization, called an SCA, was once required in order for a station to begin broadcasting a multiplex service. Although the familiar term SCA is still used, the authorization is no longer required. Broadcast licensees may begin transmitting multiplex services without prior notification of or authorization from the FCC.

Under certain conditions when SCA multiplex subcarriers are operated, a total modulation of up to 110% is legal. Figure 4.4-1 shows the baseband frequency ranges and modulation limits for various modes, from monophonic to stereophonic plus fully loaded SCA operation.

Figure 4.4-1(a) represents the basic monophonic program mode, where the baseband width is limited to approximately 15 kHz and no other signals are multiplexed. In this case all the modulating energy is contributed by the main channel. Not more than 100% modulation is permitted in this case.

Figure 4.4-1(b) shows the baseband with SCA operation in addition to monophonic main channel service. Total SCA injection up to 30% is permitted within the band from 20 kHz to 99 kHz. This injection figure may be contributed by one or more SCA subcarriers. To insure that the bandwidth of the main carrier (and its interference to other stations on adjacent and alternate channels) is not significantly increased, the arithmetic sum of all modulation must not exceed 100% plus 1/2 of the SCA injection, subject to a total modulation limit of 110%. SCA injection between 75 kHz and 99 kHz may not exceed 10% under any conditions.

Note that the Commission Rules permit transmission of multiplex subcarriers when *no* broadcast program service is carried on the main channel, provided that the above modulation rules are met.

Figure 4.4-1(c) shows the basic stereophonic sound program mode, without SCA operation. As is the case for monophonic program operation, the modulation must be limited to 100%.

Figure 4.4-1(d) adds a single band for SCA operation to the stereophonic mode. While total SCA injection may be up to 20%, no more than 20% total injection may be employed within the frequency bands from 53 kHz to 75 kHz and 10% from 75 kHz to 99 kHz. (There is one exception to determining total SCA injection involving subcarriers that are multiples of and synchronous to the stereophonic pilot. This is discussed in the SCA multiplex section of the chapter.) The modulation contributed by the main channel and stereophonic subchannel signals must be no more than 100% minus 1/2 the total SCA injection. Since the total SCA injection may be up to 20%, total modulation may be up to 110%.[4]

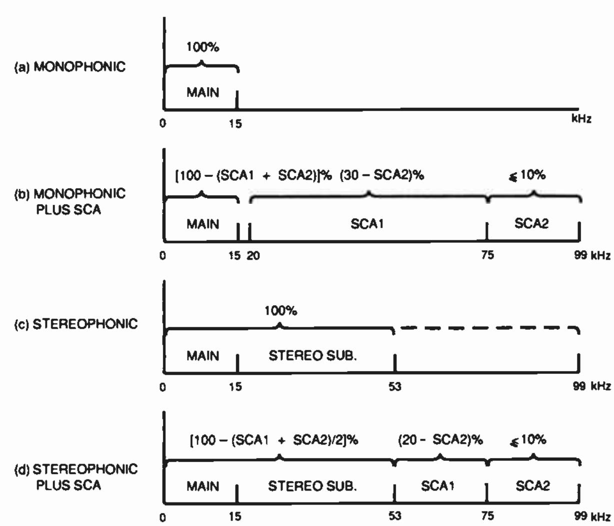


Figure 4.4-1. FM baseband scenarios with allowable modulation limits for monophonic, stereophonic, and SCA operation. Modulation percentages are referred to 75 kHz carrier deviation.

Generating The Stereophonic Baseband Signal

Figure 4.4-2 shows the composite baseband that modulates the FM carrier for biphonic broadcasting. (SCA multiplex subchannels are not part of this band and will be discussed later.)

The two channel stereo baseband has a bandwidth of 53 kHz, and consists of:

- A main channel ($L+R$) which consists of the sum of left plus right audio signals. This is the same signal broadcast by a monaural FM station, but it is reduced by approximately 10% to allow for the stereo pilot injection.
- A stereophonic sound subchannel ($L-R$) is required, consisting of a double sideband amplitude modulated subcarrier with a 38 kHz center frequency. The modulating signal is equal to the instantaneous difference of the left and right audio signals. The subcarrier is suppressed to maximize modulation capability. The pairs of AM sidebands have the same peak modulation potential as the main channel.
- A 19 kHz subcarrier *pilot* which must be exactly 1/2 the frequency of the stereophonic subcarrier and very nearly in phase to it. It supplies the reference signal needed to synchronize the decoder circuitry in receivers. The frequency tolerance of the pilot is ± 2 Hz and it must modulate the main carrier between 8 and 10%.

In general two principles have been used to generate the stereophonic subchannel—*time division multiplex* (TDM) or switching method and *frequency division multiplexing* (FDM) or matrix method.

Frequency Division Multiplexing

A basic method for generating the stereophonic baseband involves the direct generation of the double sideband suppressed $L-R$ subchannel along with the $L+R$ channel.

A simplified block diagram of the FDM system is shown in Figure 4.4-3. Both left and right audio channels are pre-emphasized and low pass filtered. In the matrix the left and right audio signals are both added and subtracted. The audio signals are added to form the $L+R$ main channel which is also used as the monaural broadcast signal.

The subtracted signals are fed to a balanced modulator which generates the $L-R$ subchannel. Since a balanced modulator is used, the carrier at 38 kHz will be suppressed, leaving only the modulated sidebands.

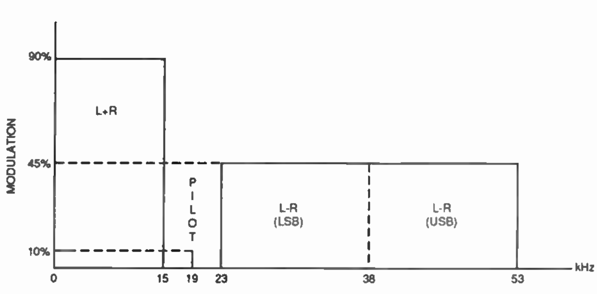


Figure 4.4-2. Biphonic (two-channel) stereo baseband.

The 38 kHz oscillator is divided by 2 to make the 19 kHz pilot tone. Finally, the main channel, stereophonic subchannel and pilot are combined in the proper proportions (45 + 45 + 10) to form the composite output.

An examination of the composite stereo waveform in the time domain, such as displayed by an oscilloscope, is helpful. First consider a 1 kHz sine wave applied equally to the L and R audio inputs. This is shown in Figure 4.4-4(a) without a pilot signal. The only frequency present in the spectrum graph is 1 kHz, since the matrix produces no difference signal necessary to generate sidebands in the stereophonic subchannel.

Figure 4.4-4(b) illustrates the ideal composite signal when the same 1 kHz tone is applied to the L and R inputs but exactly out of phase. With the pilot still off, two frequency components at 37 kHz and 39 kHz are generated. No $L+R$ signal appears from the matrix, thus only the sidebands of the modulated 38 kHz subchannel are present.

The symmetrical envelope shown represents a double-sideband suppressed carrier (DSBSC) AM signal. Note that the amplitude of each sideband is 1/2 that of the $L+R$ component in Figure 4.4-4(a). In the receiver's stereo decoder the sidebands are added together to produce an output equal to the full left signal.

Finally, consider the waveform in Figure 4.4-4(c), when the composite signal (still without pilot) is generated by applying a 1 kHz tone to the L input alone. The baseline of the waveform envelope will be a straight line if there is no amplitude or phase difference between the main channel and subchannel. Three frequency components are present: 1 kHz, 37 kHz and 39 kHz. These sidebands are each 1/2 the voltage amplitude of the 1 kHz signal in the main channel; together they equal the energy of the main channel in this instance.

The last diagram looks the same when an R-only 1 kHz tone is applied, but the phase of the two sidebands would be reversed with respect to the 38 kHz subcarrier (and the pilot). Adding the pilot at 8-10% produces similar waveforms, but oscilloscope display of the waveform baseline is fuzzier. For this reason, most stereo generators allow the pilot to be turned off for baseline measurements.

Time Division Multiplex

A different type of stereo generator is in use which produces a result similar to frequency division multiplexing by using a switching technique.

Generation of the both the $L+R$ and $L-R$ channels is accomplished by an electronic switch that is toggled by a 38 kHz signal. The switch alternately samples one audio channel and then the other, as shown in Figure 4.4-5. According to Nyquist criteria, the original signal can be reconstructed from periodic samples, provided that the samples are taken at a rate at least twice the frequency of the highest audio frequency component (approximately 15 kHz in broadcast FM).

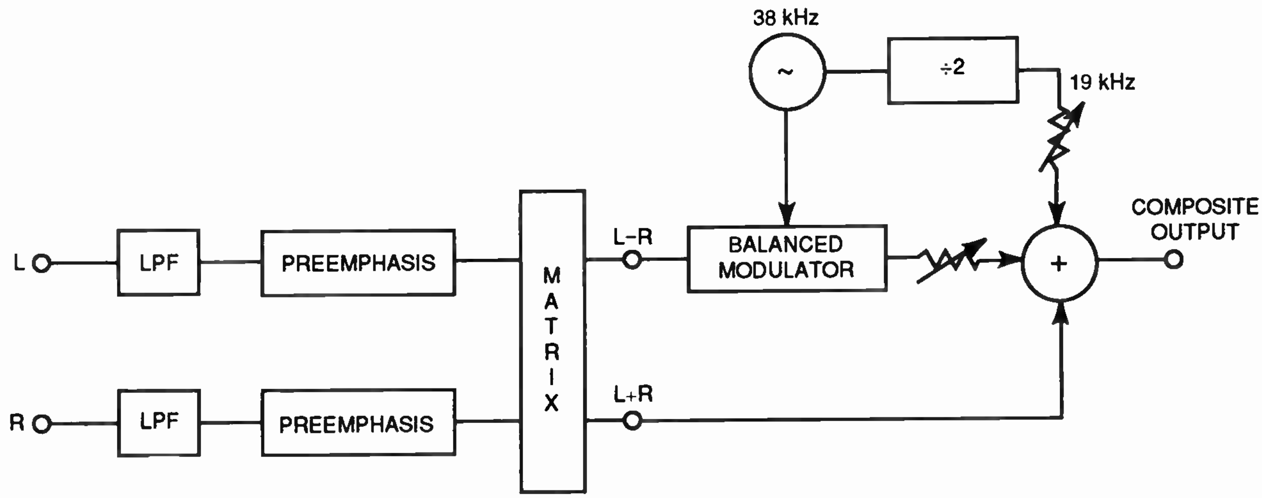


Figure 4.4-3. Functional blocks of a frequency division multiplex stereo generator.

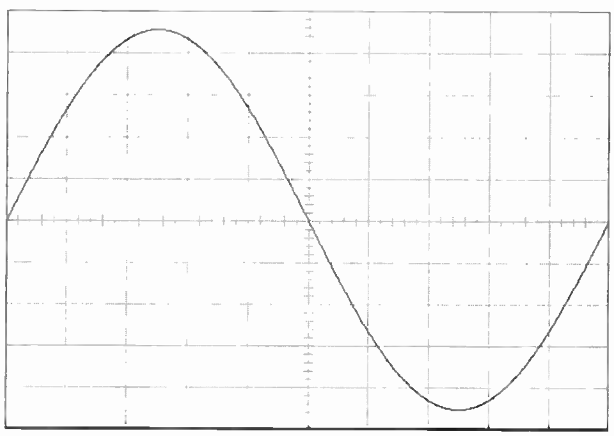


Figure 4.4-4a1. 1 kHz Left at 91%, 1 kHz Right at 91%, 19 kHz at 0%. Identical sinusoidal L and R inputs (Courtesy of Quantics.)

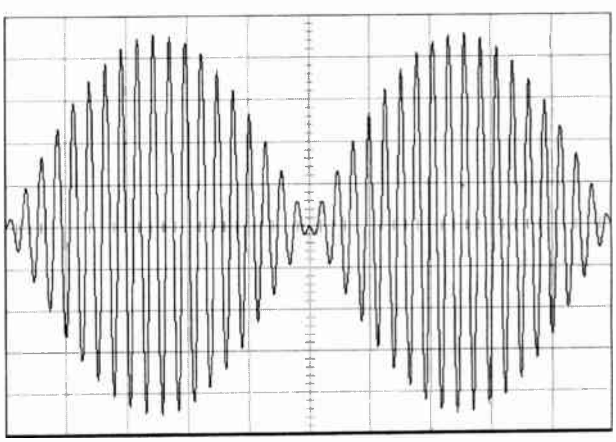


Figure 4.4-4b1. 1 kHz Left at 91%, 1 kHz Right at 91%, 19 kHz at 0%. Identical but out-of-phase sinusoidal L and R inputs.

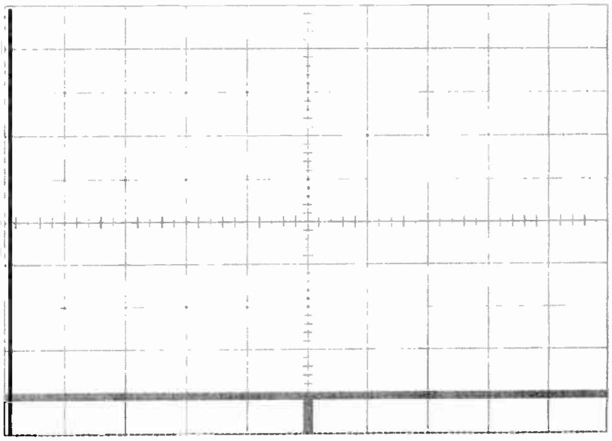


Figure 4.4-4a2. Same modulation conditions; only 1 kHz fundamental at 90% modulation (−0.92 dB) is present. (Courtesy of Quantics.)

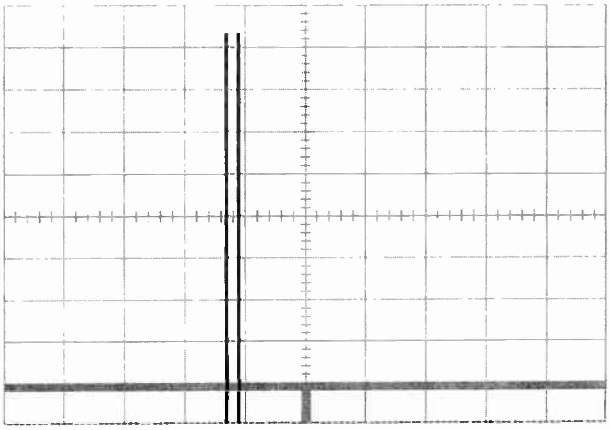


Figure 4.4-4b2. 0-100 kHz baseband spectrum; 1 kHz difference signal appears as upper and lower sidebands of 38 kHz subcarrier each at 45% modulation (−6.94 dB) for 90% total modulation. (Courtesy of Quantics.)

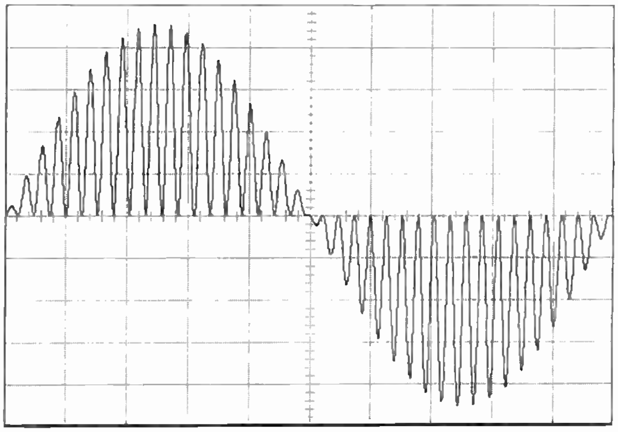


Figure 4.4-4c1. 1 kHz Left at 91%, 19 kHz at 0%. Sinusoidal L input at 90% modulation. (Courtesy of Quantics.)

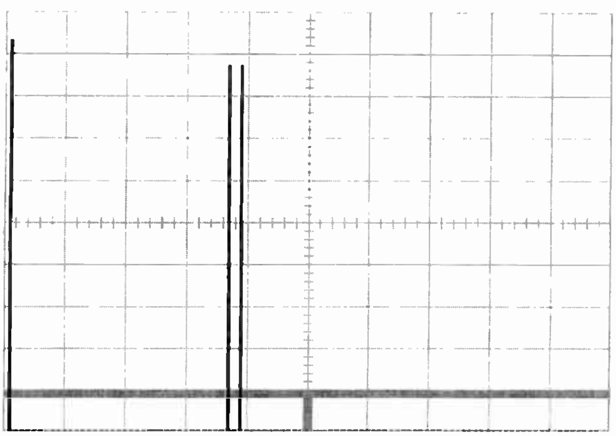


Figure 4.4-4c2. Baseband spectrum of 4c; modulation of 1 kHz signal appears a fundamental frequency at 45% level and as upper and lower sidebands of 38 kHz subcarrier at 22.5% each (−13.0 dB). (Courtesy of Quantics.)

Figure 4.4-6 shows the output waveform for the TDM generator in the time domain (as an oscilloscope would display the signal) for a sequence of input signals. The diagrams at the right of the waveform show the same signal in the frequency domain (as would be displayed on a spectrum analyzer).

In Figure 4.4-6(a), there are no input signals present. Ideally, no output signals are possible, and in practice only a small amount of leakage of the switching transients are present. Since the transfer time of the switching signal is extremely quick, harmonics of the fundamental 38 kHz are possible.

A 9 kHz audio tone is applied to the L and R inputs in Figure 4.4-6(b). The 9 kHz input signals are combined at full amplitude (90% modulation) and no subchannel sidebands are generated.

In Figure 4.4-6(c), only the left channel has a signal present. As the switch selects the L audio line, samples are passed along to the composite output. Therefore, the output waveform shows the same signal, chopped into segments of 1/38,000th of a second. Since the total area under the waveform has been divided in half, it should be apparent that the energy of the 9 kHz signal in the $L+R$ channel is only half the amplitude that it would be if an equal 9 kHz signal were also present at the right channel. The equation for the output signal e for an input signal σ at any instant t is:

$$\begin{aligned} e &= 1/2 \sin \sigma t \\ &\text{(main ch. audio)} \\ &+ 1/\pi[\sin(\phi + \sigma)t + \sin(\phi - \sigma)t] \\ &\text{(DSBSC)} \\ &- 1/3\pi[\sin(3\phi + \sigma)t + \sin(3\phi - \sigma)t] \\ &\text{(3rd harmonic)} \\ &\ldots \text{etc. (higher harmonics)} \end{aligned} \qquad [1]$$

Figure 4.4-6(c) shows the original 9 kHz signal (at half amplitude), and a pair of sidebands centered about the 38 kHz switching frequency. No 38 kHz signal

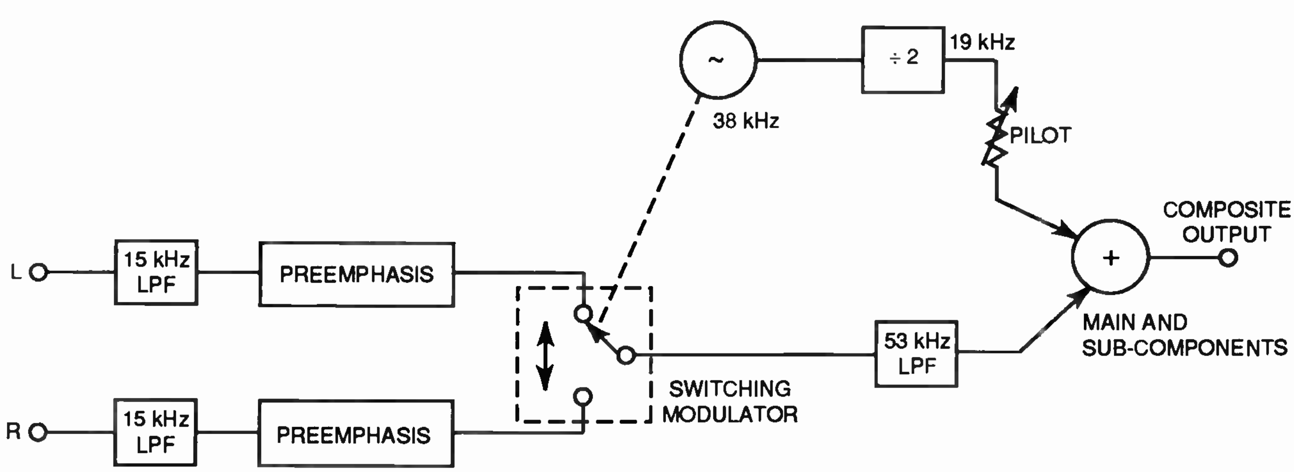


Figure 4.4-5. Functional blocks of a time division multiplex stereo generator.

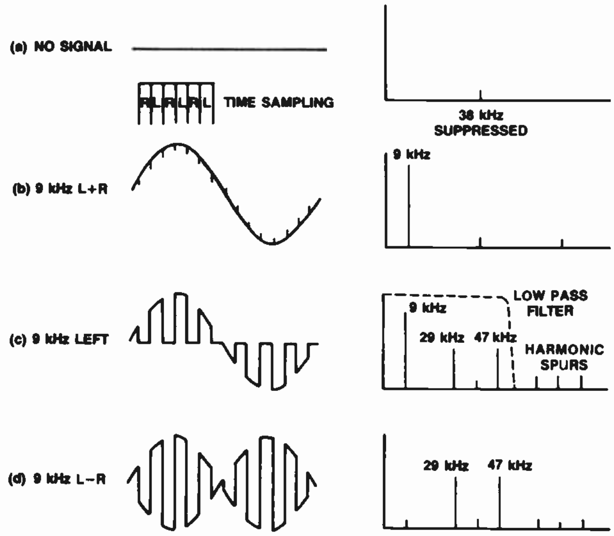


Figure 4.4-6. Time domain and frequency domain diagrams of stereo baseband signals.

is generated if the switching waveform has perfect symmetry, that is, if the switch is connected to the left and right channels for precisely equal periods. Note that a harmonic of the stereophonic subcarrier is shown, centered around 114 kHz which is three times the switching frequency. Only one extra term was shown in the equation; however, other terms at the 5th, 7th are present.

In addition to the odd order harmonics of the 38 kHz subchannel, asymmetry in the switching signal or other circuit imbalances can create some sidebands centered about the second harmonic at 76 kHz. All these harmonics must be removed by filtering, as shown in the diagram.

When the odd harmonics are filtered out the proper DSBSC waveform results. However, it is slightly greater in amplitude than the $L+R$ signal because the fundamental component of the square wave is $4/\pi$ or 1.27 times larger than the square wave amplitude. This is easily corrected by adding enough of the L and R audio to the output to equalize the amplitude.

In Figure 4.4-6d, the TDM signal is shown when the L and R signals are equal in amplitude and exactly reversed in phase. This waveform matches the composite stereo signal shown in Figure 4.4-4b.

The composite lowpass filter must have very steep cutoff characteristics but should have flat amplitude response and linear phase shift with frequency (equal time delay at all frequencies) below 53 kHz. While this approach to stereophonic generation is simple and stable, the filter can degrade stereo separation, especially at higher audio frequencies.

The 19 kHz pilot squarewave from the ÷2 digital divider must also be filtered to remove harmonics. This additional time delay (phase shift) of the pilot with respect to the 38 kHz information must be compensated to have optimum channel separation.

A significant improvement on the original switching concept is shown in Figure 4.4-7. As mentioned earlier, the higher order terms of the square wave-driven switch are responsible for generating the harmonics of the 38 kHz subchannel which must be removed by filtering. By using a *soft switch* to connect back and forth between the L and R channels it is possible to eliminate the lowpass filter and its side effects.

This is accomplished by using the electrical equivalent of a variable attenuator, shown in the diagram by a potentiometer. The slider is driven from end to end of the potentiometer by a sinewave. Since a sinewave is represented only by a single, fundamental frequency, the signal output at the slider has the proper DSBSC characteristics without the harmonics. The equation for the composite signal generated in this way is:

$$e = \underbrace{1/2 \sin \sigma t}_{(L+R\ audio)} + \underbrace{1/\pi[\sin(\phi + \sigma)t + \sin(\phi - \sigma)t]}_{(38\ kHz\ DSBSC)} \qquad [2]$$

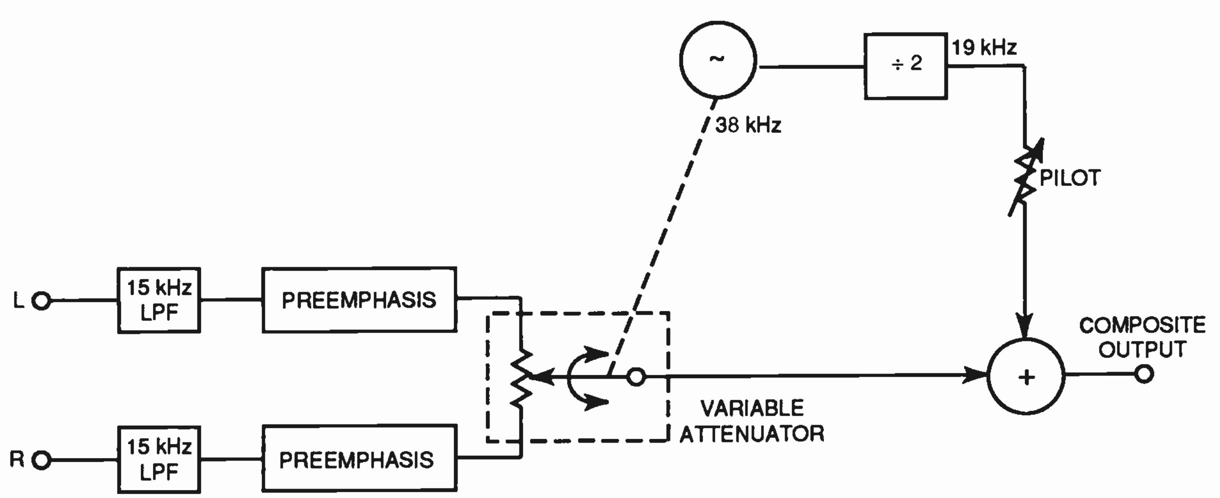


Figure 4.4-7. Functional blocks of a time division multiplex stereo generator using a variable attenuator.

As the equation shows, only the fundamental sidebands of 38 kHz are present in the sampled signal, along with the main channel component. Like the fast switching TDM system, the $L+R$ and $L-R$ channels are generated in one operation so that the circuit remains relatively simple. No filter of the output is required, provided that the 38 kHz sinewave is free from harmonics and the variable attenuator has good linearity.

Digital Stereo Generation

The major manufacturers of stereo generators have moved to all digital composite generator designs. These units carry out the same functions as analog stereo generators, but with the flexibility and consistency commonly expected of digital audio systems. Some generators also perform all the audio compression and peak limiting functions in the digital domain. This chapter is confined to a discussion of digital stereo generation techniques.

Digital implementation is made easier if the output sampling rate is a binary multiple of either of the two steady state vectors in the composite signal: the stereo pilot or the $L-R$ subcarrier. For example, an oversampling rate of 304 kHz is sometimes employed (this is eight times the subcarrier frequency, and 16 times the pilot).

The input sampling rate is also an important consideration. The reader will note that switching type stereo generators, discussed earlier, operate at only 38 kHz. While higher rates do not yield any additional improvement in fidelity, it is desirable to use a higher rate throughout the generator system.

While Digital Compact Discs© are fixed at 16-bit quantization precision, most digital stereo generator designs use more bits per sample. This is because analog-to-digital (A/D) conversion devices do not perform to the theoretical limits of 16 bits, and because the mathematical precision of the processing architecture can produce rounding errors that contribute noise and distortion to the processed signal. Eighteen to 24 bit data paths are generally chosen.

Analog audio for each channel is passed through an anti-aliasing lowpass filter and is then A/D converted into digital data streams. At this point, generators with audio processing carry out algorithms that gain-control, pre-emphasize, limit and filter the digital audio stream. The dynamic effects are very similar to analog processors even though the treatment is entirely digital.

The processed L and R signals are then numerically matrixed using simple addition into $L+R$ and $L-R$ channels. The digitized 38 kHz sinewave is derived from a look up table and digitally multiplied by the $L-R$ channel to produce the 38 kHz DSBSC. The 19 kHz pilot is generated by another look up table that is locked via software control so that the phasing relative to the 38 kHz subchannel is perfect. The pilot and 38 kHz subchannel are finally summed and applied to a digital-to-analog (D/A) converter to form the complete composite stereo signal. Some manufacturers provide a separate digital output port for digital FM exciters.

Stereo Decoder Circuits

Stereo FM receivers include a circuit to convert the multiplexed composite signal at the FM detector into the original left and right audio channels transmitted by the FM station. There are at least as many ways to decode the stereophonic signal as there are ways to encode (generate) the composite signal. In practice, only one type of decoder is commonly used, the *phase-locked loop* (PLL) integrated circuit.

The circuit in Figure 4.4-8a is seldom used, but is shown for comparison. It is the closest complement to stereo generators using frequency division multiplexing. At the input, the composite signal is split with three equal time-delay filters into the main $(L+R)$ channel, pilot signal and stereo $(L-R)$ subchannel. Next, the pilot is doubled to 38 kHz and this regenerated carrier is reinserted into the double-sideband AM signal from the subchannel filter. This AM signal is demodulated to yield the $L-R$ (difference) audio. Finally, the $L+R$ and $L-R$ signals are combined in a sum and difference matrix to produce the original L and R audio channels.

Because of the costly filters needed to separate the composite spectrum, the frequency division multiplex circuit is not used in consumer equipment. Similar circuits have been used in broadcast modulation monitors, where metering of the separate channels is required.

The circuit shown in Figure 4.4-8b is universally used, due to its simplicity, high performance and low cost. While this stereo decoder is commonly referred to by its internal PLL, its performance is really distinguished by its time division demultiplexer (shown in the dashed box as a toggle switch).

Following a buffer amplifier, the composite baseband signal is sampled by a PLL within the IC. A voltage controlled oscillator, usually running at 76 kHz (four times the pilot frequency) is held in phase with the pilot by a reference signal from the phase comparator and loop filter. When divided by two, the result is a square wave at 38 kHz having nearly perfect duty cycle (high and low states have equal timing) and very fast rise and fall times. This signal is ideal for driving the output audio switcher (demultiplexer). This stage is a transistor matrix designed to rapidly transfer the composite baseband to the L and R audio output in time with the switch in the station's stereo generator. Fast, clean audio switches are relatively easy to make and do not drift.

Because PLL stereo decoders normally use square wave switching, the circuit is able to demodulate baseband signals that are odd harmonics of 38 kHz. The third harmonic (114 kHz) is most troublesome, since noise and spurious signals near this frequency are shifted to the audio baseband, as is the 38 kHz stereophonic subchannel. Engineers should be watchful of the frequency band centered on 114 kHz in their transmitted signal since audible noise may occur in consumer receivers.

Recent stereo decoder designs have reduced sensitivity to energy outside the composite baseband. One

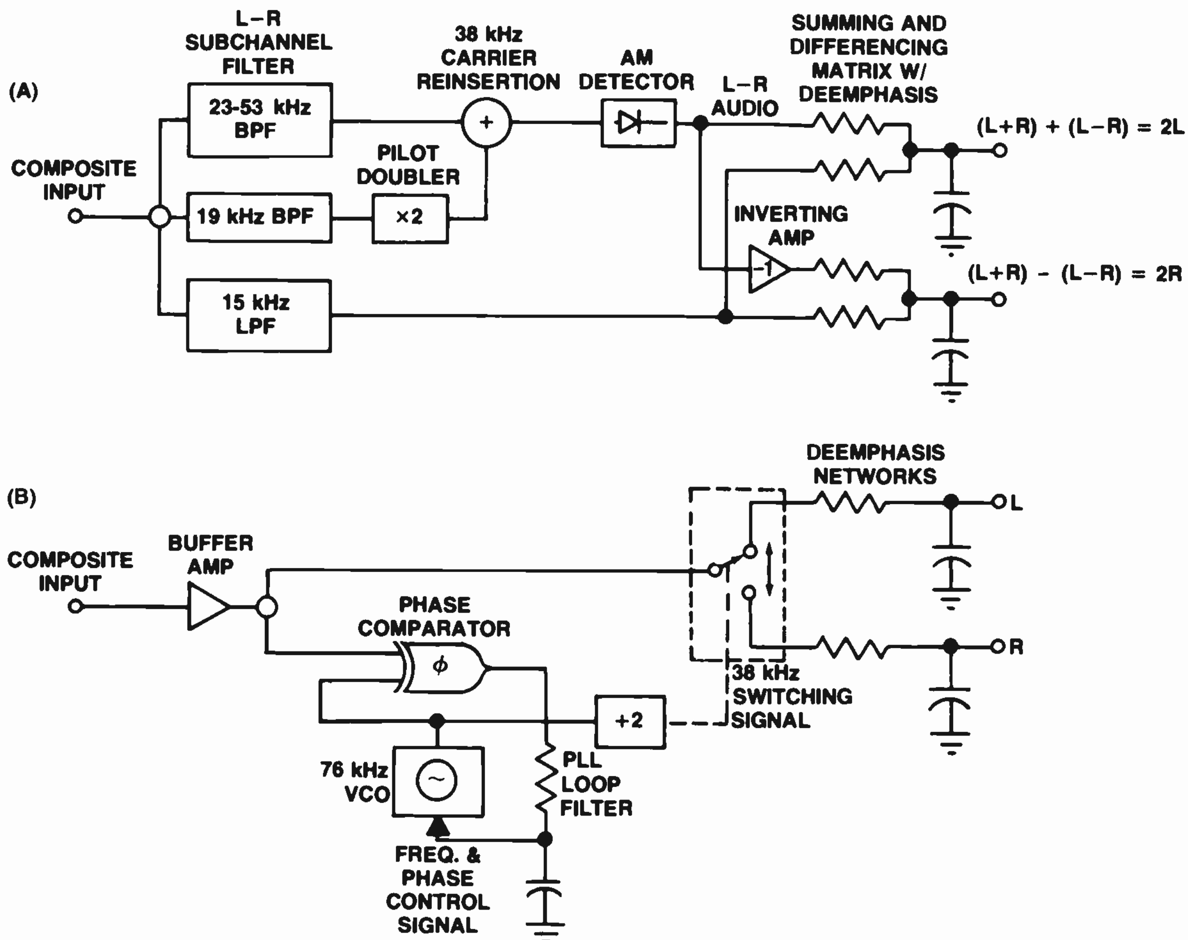


Figure 4.4-8. Functional blocks of stereo decoders using (A) $L+R$ and $L-R$ matrixing and (B) phase-locked time-division multiplexing.

approach utilizes a second composite audio toggle switch operated at 114 kHz. The demodulated product is inverted and mixed equally with the 38 kHz switching outputs, canceling the response to signals in the 114 kHz range. The other approach—Walsh demodulator—applies a properly timed stair step imitation of a sinusoid to a digital multiplier in the output signal path.[5] This reduces sensitivity to 3rd and 5th harmonics, as well as adjacent channel noise and interference, by up to 20 dB.

FM SCA TRANSMISSION

From its beginning as a broadcast service, people have recognized the potential of FM for multiplexed services. As early as 1940, the FCC permitted multiplex facsimile transmission on FM stations, but not until much later did auxiliary FM services attain a wide acceptance among FM broadcasters.

In 1955, the FCC established the SCA and created an entirely new industry. The original intent of the authorization was to permit programming of background music to offices, stores, restaurants, etc., where it was uneconomical to provide this service via telephone lines. For many commercial FM stations, the SCA operation became a major source of revenue which enabled them to survive in the 1950s.

By the early 1980s, improvements in transmitter and receiver technology and the desire for new revenue prompted commercial and noncommercial broadcasters to seek changes in the SCA rules. In a series of rulemakings in 1982 and 1983 the Commission made numerous changes to expand technical opportunities and reduce legal regulation.[6] These changes extended the baseband frequency limit from 75 kHz to 99 kHz, allowed any type of subcarrier modulation to be used, changed the subcarrier injection requirements to permit multiple services and increased limits for the total modulation during SCA multiplex operation to reduce main channel modulation loss.

FCC Rule Requirements For SCA Operation

Section 73.319 of the FCC Rules sets forth the technical standards for FM multiplex subcarriers. However, the Commission does not set standards for mini-

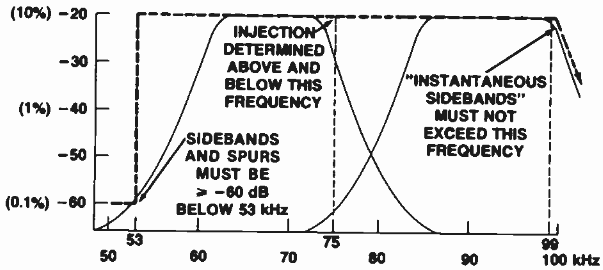


Figure 4.4-9. Injection, channel bandwidths and spurs limits for SCA operation when stereo is transmitted. Two possible subcarriers (at 67 and 92 kHz) are shown.

mum SCA subcarrier performance. This is left for the broadcaster or lessee of the service to determine. In its rules, the Commission defines the transmission conditions under which subcarriers may be operated, to minimize interference to the main channel and stereophonic subchannel and to other FM stations.

In this chapter, subcarrier injection and bandwidth are considered here only with stereophonic operation, since this is by far the most common FM mode. Figure 4.4-9 shows the upper portion of the composite baseband, from approximately 50 kHz to 100 kHz. This is the same as the upper portion of Figure 4.4-2 with the addition of two hypothetical subcarriers centered at 67 kHz and 92 kHz. Any number of SCA subcarriers may be operated in this frequency range provided that total bandwidth and injection limits are met.

On the baseline are several frequency markers at the upper and lower frequency limits of the two subcarrier ranges. The actual frequencies of the subcarriers are not specified.

Alongside each subcarrier are arrows marking the level in dB below 100% modulation of the main carrier. For example, -20 dB marks the injection at the center frequency of both subcarriers. Since 10% is the maximum injection permitted under the Commission rules within each SCA subchannel.

$$\text{Injection} = 20 \times \log_{10}(0.1) = 20(-1) = -20 \text{ dB} \quad [3]$$

At 53 kHz, an arrow marks a level of -60dB. The FCC requires that any frequency modulation of the main carrier due to the SCA operation shall be at least 60 dB below 100% modulation in the frequency range of 50 Hz to 53 kHz when stereo is transmitted. This figure must include spurious and intermodulation products as well as subcarrier sideband energy.

At 99 kHz, the level of -20 dB is marked, denoting the FCC requirement that "instantaneous sidebands" be restricted within this frequency limit. The Commission has not officially defined instantaneous sidebands, but it is normally considered to be the instantaneous frequency of the subcarrier at its peak deviation (for frequency modulated subcarriers) or the highest sideband frequency (for amplitude modulated subcarriers).

Interference Between FM SCA Subcarriers

In practice, two subcarriers should be separated as far apart in frequency as possible, while observing the limit of spurious energy below 53 kHz and the instantaneous sidebands at 99 kHz. While the FCC Rules are silent on the choice of SCA frequencies, 67 kHz and 92 kHz have become the defacto standards for FM subcarriers. The first frequency was adopted when the original stereophonic standards restricted the subcarrier spectrum between 53 kHz and 75 kHz. The second frequency was recommended to situate the instantaneous sidebands below 99 kHz while maintaining a safe separation from a 67 kHz subcarrier.

Some overlap of the subcarrier sidebands does occur, as depicted in Figure 4.4-10. This does not cause significant interference between the two SCA subchannels when the systems use frequency modulation. Following is a list of bench test results of a standard table model SCA receiver:

- Unweighted 67 kHz SCA S/N ratio at 65 dBf, referred to 5 kHz deviation: 57.5 dB.
- Same, with 4 kHz tone modulation of 92 kHz subcarrier at 7 kHz deviation: 56.5.
- Same as first case with tone modulation of main channel at 67.5 deviation: 52.5.

The above data shows that crosstalk is greater from main-to-67 kHz subcarrier than from a 92 kHz subcarrier into the 67 kHz subchannel demodulator.

Other Types of Analog Subcarrier Modulation

Other frequencies and total number of subchannels are permissible, according to the occupied bandwidths and interference margins required. A hypothetical SCA baseband spectrum combining five amplitude-modulated subcarriers is shown in Figure 4.4-11. A variety of bandwidths and injection levels are used.

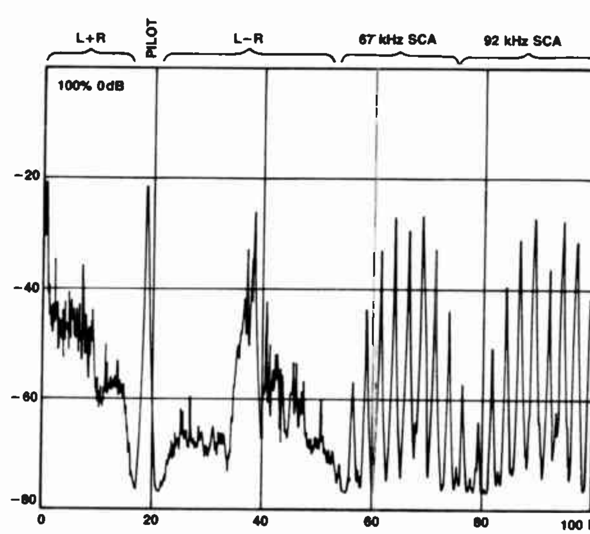


Figure 4.4-10. Composite baseband of FM station with 67 kHz and 92 kHz FM-SCAs, modulated with 2.5 kHz tone at 5 kHz and 7 kHz peak deviation, respectively. Note overlap of sidebands between subcarriers. Station is carrying stereo programming.

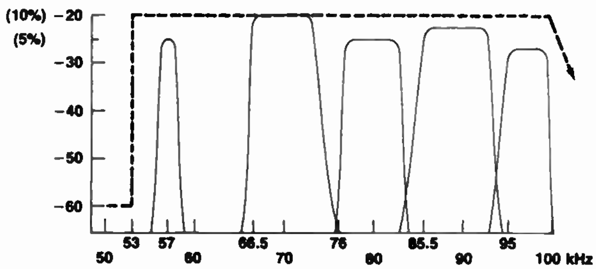


Figure 4.4-11. Carrier envelope spectrum, DSB-AM and SSB-AM SCA subcarriers. Vertical markings are dB referred to 75 kHz deviation.

In Figure 4.4-11 note that single sideband suppressed carrier (SSBSC) AM systems are used for the four other subchannels. On the first (at 66.5 kHz), 10% peak envelope injection is being used, while the other three use levels which bring the total to 20%.

At 57 kHz, the envelope of a narrow subcarrier is shown. Actually, standards exist to interleave two subcarriers within the same channel. One is the *Radio Data System* (RDS) which is in use by a large number of Western European countries for personal paging, station and program identification, emergency alerting, etc. This 57 kHz system is currently used by over 700 stations in the U.S. The modulated carrier spectrum resembles suppressed-carrier double sideband AM that is locked to the third harmonic of the pilot. The digital data rate is 1187.5 bits/sec and nominal deviation of the carrier is 2.67% (±2.0 kHz).

Automotive Road Information (ARI) is another 57 kHz system developed in Europe. It is used to identify stations that broadcast traffic information and to activate specially equipped FM car radios during traffic announcements. ARI is a slow rate signaling system with very narrow bandwidth. The 57 kHz ARI carrier is centered within the suppressed carrier region of the RDS signal (if RDS is transmitted).

Examples Of Analog SCA Operation

Figure 4.4-12a shows the functional blocks of a standard aural SCA generator using frequency modulation. At the input, audio is lowpass filtered to limit the modulation sidebands that could interfere with the stereophonic subchannel or an upper SCA subchannel.

A 5 kHz audio cutoff is frequently used so that fidelity compares favorably to unequalized telephone service. Using a Bessel function analysis one can determine the peak deviation which may be used to maintain FM sidebands below 53 kHz within FCC limits:

Sinusoidal SCA Modulating Frequency	Maximum SCA Peak Deviation
3.5 kHz	5.0 kHz
5.0	3.5
7.5	2.0

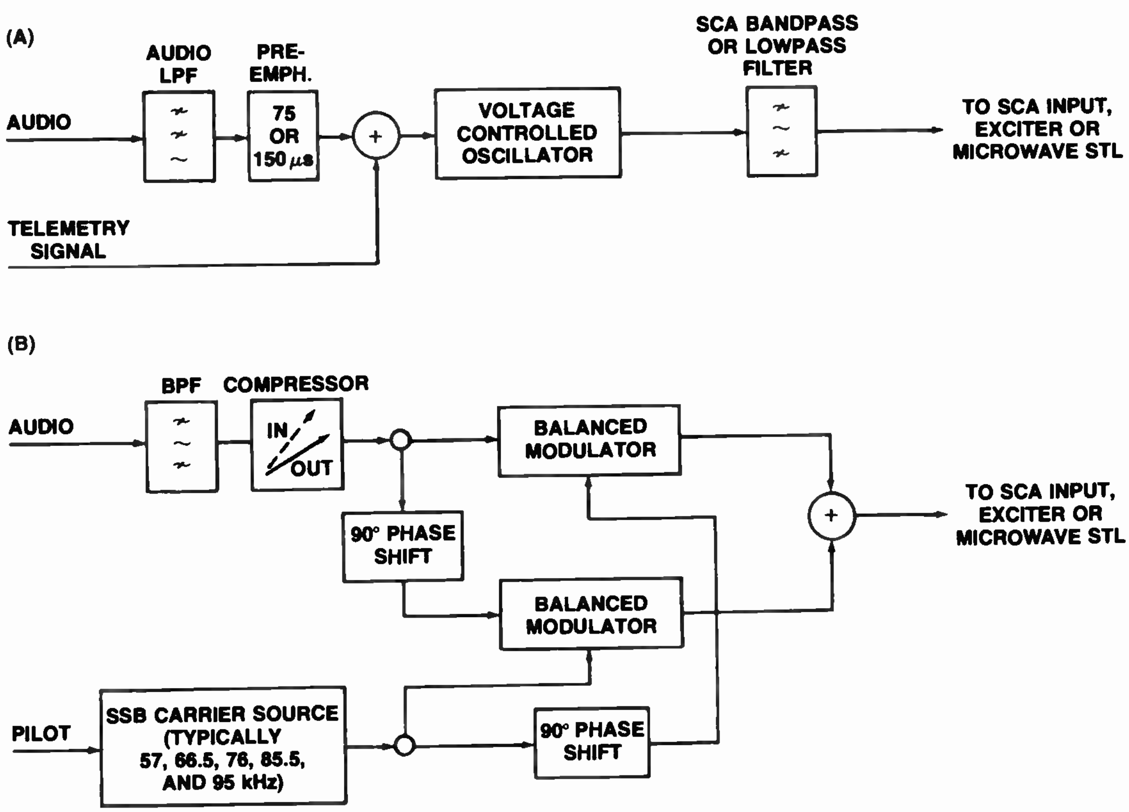


Figure 4.4-12. Functional blocks of (A) a frequency-modulated SCA subcarrier generator, and (B) a single-sideband AM subcarrier generator.

For the protection requirements, equipment manufacturers often recommend 3.5 or 4.0 kHz peak deviation with a 5 kHz lowpass filter when stereo is transmitted. In practice, slightly higher deviation may be used since the calculations assume single tone modulation which has greater modulation power than program audio.

Pre-emphasis follows the lowpass filter in the sample FM SCA generator. A value of 150 μsec is commonly used to combat the rising high frequency noise characteristic in the FM SCA channel. Unfortunately, this extreme boost (almost 17 dB at 5 kHz, compared to 400 Hz) requires substantial amounts of peak limiting to allow full modulation at low and middle audio frequencies.

The processed audio signal is fed to the FM modulator, usually with a provision for a subaudible (less than 30 Hz) channel for telemetering return. The telemetry channel takes advantage of the extended low frequency response capability of FM subcarriers by sending tone signals in the 10 to 30 Hz range at a level of 15 to 25 dB below reference peak deviation. A highpass filter is inserted ahead of the audio input when subaudible telemetry is used to avoid interference to the low frequency channel.[8]

Ideally, no filtering of the frequency modulated subchannel is required. However, some of the FM generator circuits produce spurious or harmonic energy. For this reason, a bandpass or lowpass filter is used that passes all the significant sidebands but suppresses spurious signals to acceptable levels. The modulated subchannel signal is amplified to a level required by the subcarrier input of the FM exciter or studio transmitter link equipment.

A simplified block diagram of a SSBSC AM subchannel generator is shown in Figure 4.4-12b. The input audio is bandpass filtered to control subchannel bandwidth (to 5 kHz, for example), and may be compressed according to specific amplitude and spectral standards. A complementary expander circuit in the receiver restores the original program dynamics and reduces noise in the channel. Large amounts of pre-emphasis are not as important since the spectral characteristic of the noise is flat with a linear modulation system, not rising with frequency as in angular modulation systems.

The processed audio signal is fed to a pair of balanced modulators, one path going through an all-pass filter which shifts the audio phase 90° throughout the required audio range. The balanced modulators are driven by a subcarrier frequency source which may be frequency locked to a harmonic of the stereophonic pilot (and its half frequency of 9.5 kHz, as shown). One of the modulators is driven by a carrier signal which is shifted 90°. When the two modulated carriers are summed, cancellation of the lower (or upper) sidebands creates the desired single sideband signal.

Since SSBSC AM requires a carrier for detection, the pilot provides an excellent source for reinsertion at the receiver. With monophonic audio no pilot is transmitted, so another carrier reference must be used. One approach is to carry the 19 kHz reference at a reduced level which will not trigger the stereophonic decoders in FM receivers.

SCA DATA SYSTEMS

SCA subchannels have characteristics that are quite favorable for transmitting data. While the channels are noisy for high fidelity audio services, they can provide good to excellent data error performance, and moderately fast data rates are possible without complicated or expensive receivers. Combined with the high power and elevation of FM broadcast stations, SCA data systems may be operated over large service areas.

There are many methods of putting data on an SCA subchannel, the simplest being *audio frequency shift keying* (AFSK)—the connection of a telephone type data modem to the audio terminals of the SCA generator.

The cost and complexity of this form of SCA data broadcasting is low. However, this technique occupies the bandwidth of a standard audio subcarrier and the highest practical speed is about 2,400 bps. This may be acceptable where the hardware must be off-the-shelf and data capacity of baseband spectrum is not a consideration.

Direct Data Modulation

Direct data modulation eliminates the intermediate steps that convert the data into audio signals, allowing the subcarrier to be directly modulated by the binary data. This efficiency allows a greater amount of data to transmitted within the available SCA subchannel bandwidth.

In *non-return-to-zero frequency shift keying* (NRZ/FSK), the two binary states are represented by two different subcarrier frequencies. For example, a "0" could be 65 kHz and a "1" 69 kHz. FSK is the most popular form of direct data modulation at this writing, probably because of its robust performance, simplicity and low cost. Using careful design, a modulated bandwidth similar to that of background music services (about ±16 kHz at −25 dB below subcarrier injection) and a maximum speed of 19 kbps are currently possible.[9]

Data FSK systems must use filtering methods to control occupied bandwidth, since the rapid frequency transitions generate high order sidebands that could interfere with adjacent subchannels. Filtering is accomplished by either or both of the following methods shown in Figure 4.4-13:

- Where a pair of oscillators are alternately toggled by the data states, filtering must be done to the output spectrum to directly reduce occupied bandwidth.
- Where analog FM modulators are used, filtering the slew rate of the switching signal tends to reduce high order sidebands indirectly.

The spectrum of the composite baseband of a FM station is shown in Figure 4.4-14, carrying a 19 kbps

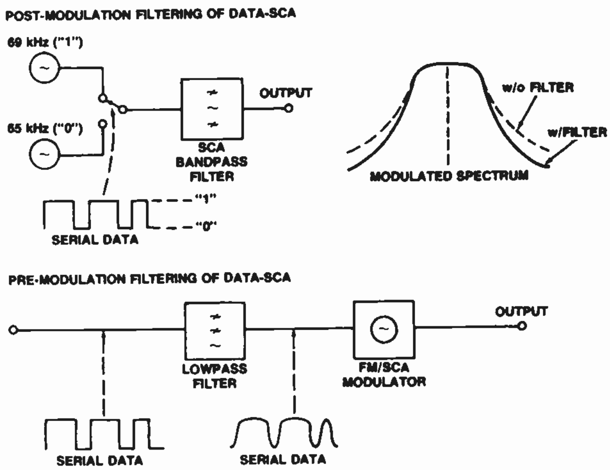


Figure 4.4-13. Two basic methods for generating and controlling the bandwidth of data subcarriers used for SCA.

data signal. The spectrum was generated by a specially designed SCA generator employing both pre- and post-modulation filtering.

Phase Shift Keying (PSK) is another technique for transmitting high speed data. With PSK, one phase of the carrier represents one binary state and a second phase (usually 180° apart) is used for the second state. High speed data transmission is possible, up to 56,000 bps, although most of the 53 to 99 kHz bandwidth might be required and receiver/decoders would be much more expensive than those for moderate speed FSK systems.

Radio Broadcast Data System

The United States Radio Broadcast Data System (RBDS) Standard was first adopted by the National Radio Systems Committee in 1993. A revised edition of this standard was adopted in 1998. The RBDS Standard uses an FM data subcarrier at 57 kHz in the FM baseband to transmit information like call sign, format type, song title, artist, and emergency information to RBDS-equipped receivers. This standard is explored in more detail in Chapter 4.6, *Radio Data Broadcasting*.

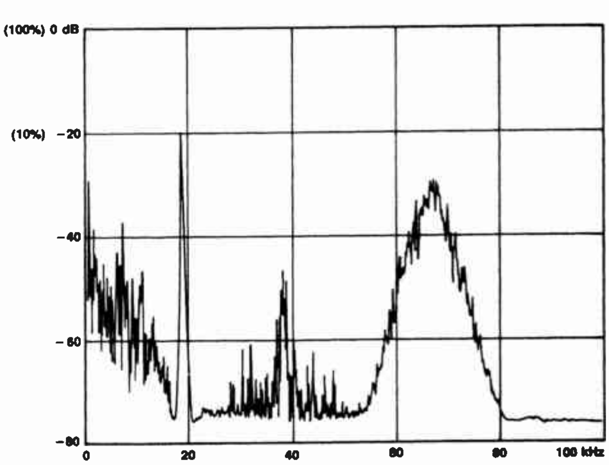


Figure 4.4-14. Composite baseband of an FM station carrying a 19 kbps frequency-shift keyed data subcarrier. Center frequency of the subcarrier is 67 kHz, with an injection of 9%. Station is carrying stereo programming.

MONOPHONIC, STEREOPHONIC & SCA PERFORMANCE

While monophonic, stereophonic and various SCA services are conveyed by the same FM carrier, performance varies considerably due to the relative carrier deviation of each and the particular frequency span occupied within the composite baseband. The performance of a channel may be expressed in terms of demodulated S/N ratio for a given radio frequency carrier level, or, conversely, coverage range for a given signal to noise ratio.

The performance of a conventional FM receiver in the presence of random (thermal) noise is commonly judged on the basis of the variation of the output S/N power ratio as a function of the carrier to noise power ratio contained within the receiver bandwidth, usually determined by the IF bandwidth. An estimate of the IF bandwidth required for transmission of (monophonic) broadcast FM is given by the following equation:[10]

$$\beta_{IF} = 2(F_{dev} + 2f_m) \qquad [4]$$

Where:

F_{dev} = the peak frequency deviation (75 kHz), and
f_m = the highest baseband modulating frequency (15 kHz).

This β_{IF} figure is rounded to 200 kHz for later use.

The equivalent input noise power in a bandwidth β launched into a noise free (cool) receiver is:

$$\begin{aligned} rB &= 4 \times 10^{-21}\beta f_N \\ &= 8 \times 10^{-16} f_N \text{ W} \end{aligned} \qquad [5]$$

Where:

f_N = the receiver noise figure, and
β = nominal noise bandwidth in Hz.

The carrier power is:

$$C = (V_{\mu V^2}/R) \times 10^{-2} \text{ W} \qquad [6]$$

Where:

$V_{\mu V^2}$ = the rms input voltage in microvolts
R = the receiver r.f. input resistance.

For R = 75 Ω the CNR in decibels is:

$$\begin{aligned} CNR &= 10 \log_{10}(C/rB) \\ &= 12.2 + 20 \log_{10} V_{\mu V} - 10 \log_{10} f_N \end{aligned} \qquad [7]$$

For example, for a 10 dB noise figure the carrier to noise ratio is 0 dB for an input voltage of 0.77 μV over 75 Ω (1.55 μV over 300 Ω).

For a half-wave dipole antenna matched to a 75 Ω receiver the input voltage is related to the field strength as:

$$V_{\mu V} = 48.5E_{\mu V}/f_{MHz} \qquad [8]$$

Where:

$E_{\mu V}$ = the field strength in μVμ
f_{MHz} = the carrier frequency in MHz.[11]

Thus, for 98 MHz (middle of the FM band)

$$V_{\mu V} = \frac{E_{\mu V}}{2} \qquad [9]$$

With the FCC F(50,50) field strength curves it is possible to relate μV to the distance to the transmitter, effective radiated power and antenna heights. Ultimately, it is possible to predict the CNR for a given broadcast system. For a 200 kHz nominal bandwidth the CNR in dB can be shown to be:

$$\begin{aligned} \text{CNR} = E_0 + P + G - F \\ + 20 \log_{10}(h/30 \times 98/f) + 6 \end{aligned} \qquad [10]$$

where

E_0 = field strength for 1 kW ERP 1000′ HAAT, in dB μV/m
P = ERP in dBkW
G = receiver antenna gain relative to dipole, in dB
h = receiving antenna height above ground, in ft
F = receiver noise figure in dB
f = frequency in MHz.

For example, if $F = 10$, $P = 10$, $G = 0$, $h = 30$, and $f = 98$, then CNR $= E_0 + 6$.

Next, the S/N ratio is established by the CNR:

$$\text{SNR} = 40 + 10 \log_{10} \text{CNR dB} \qquad [11]$$

This expression assumes a peak deviation of 75 kHz, a de-emphasis corner frequency of 2.1 kHz (75 μS) and a nominal intermediate frequency noise bandwidth of 200 kHz.[12]

Using the previous equation, a carrier to noise ratio of 0 dB would result in a signal to noise ratio of 40 dB. However, standard FM demodulators typically exhibit a threshold at about 12 dB CNR, below which the noise floor changes from a hiss to a raspy kind of noise with noticeable clicks jutting above the noise. For this reason, the monophonic channel actually deteriorates below a signal to noise ratio of about 52 dB.

A well known effect of FM demodulation is that the noise voltage in a narrow band (for example, 1 Hz) increases with frequency. This has important implications for systems which employ subchannels since both the bandwidth and center frequency affect the S/N ratios. For example, the stereophonic $(L-R)$ subchannel shown in Figure 4.4-2 extends nominally from 23 kHz to 53 kHz and occupies twice the bandwidth of the main $(L+R)$ channel. When the noise of the $L-R$ subchannel is combined during stereo decoding, the unweighted noise floor is 23.1 dB higher than the monophonic system under the same conditions.[13]

The S/N ratio of an SCA subcarrier, relative to the main channel, is a function of a number of variables and is:

$$N_S/N_O = (2f_c^2/m^2D^2)\frac{f_1^3(b_1 - \arctan b_1)}{f_0^3(b_0 - \arctan b_0)} \qquad [12]$$

Where:

f_c = subcarrier frequency
m = injection level as a percent of 75 kHz
D = deviation of FM subcarrier
f_0 = −3 dB freq. of de-emphasis in main audio channel (2.12 kHz)
f_1 = −3 dB freq. of de-emphasis in SCA audio channel (1.06 kHz)
b_0 = main channel audio output bandwidth (15 kHz)/f_0
b_1 = SCA channel audio output bandwidth (5 kHz)/f_1

Following is an example of the excess noise in the 5 kHz SCA audio channel, compared to the 15 kHz main channel, if $m = 0.1$ (10%). Note that the noise level of a 67 kHz subcarrier with 5 kHz deviation and a 92 kHz subcarrier with 7 kHz deviation are nearly equal. Comparative tests by the author have indicated that susceptibility to impulse noise (gasoline engine ignition radiation) is greater at the higher subcarrier frequency even though Gaussian noise levels are approximately the same. This may be due to a higher noise threshold in the 92 kHz demodulator (due to a wider pre-detection bandwidth than the 67 kHz demodulator) or to differences in the spectral distribution of impulse noise compared to Gaussian noise in the IF (composite) demodulator.[14]

f_c (SCA freq., kHz)	67	92	92
D (deviation, kHz)	5	5	7
N_S/N_o (above main, dB)	36.4	39.2	36.2

Figure 4.4-15 graphs the relative subcarrier noise level against increasing frequency of a subcarrier, compared to the main (monophonic) channel for the same receive conditions. Shown are two peak deviations (5 kHz and 7 kHz) for FM SCA, and SSBSC AM SCA with 10% peak envelope injection. The graph segment representing 7 kHz deviation is dashed below 70 kHz since excessive interference to the $L-R$ subchannel is possible in this frequency range when stereo is transmitted.

Single Sideband AM Subcarriers

The conditions under which standard SCA subcarriers are operated are quite close to those existing with narrow band frequency modulation. For example, when a maximum audio bandwidth f_m of 5 kHz and peak deviation f_d of 5 kHz are used, the modulation index is 1.

$$i = f_d/f_m = 5/5 = 1 \qquad [13]$$

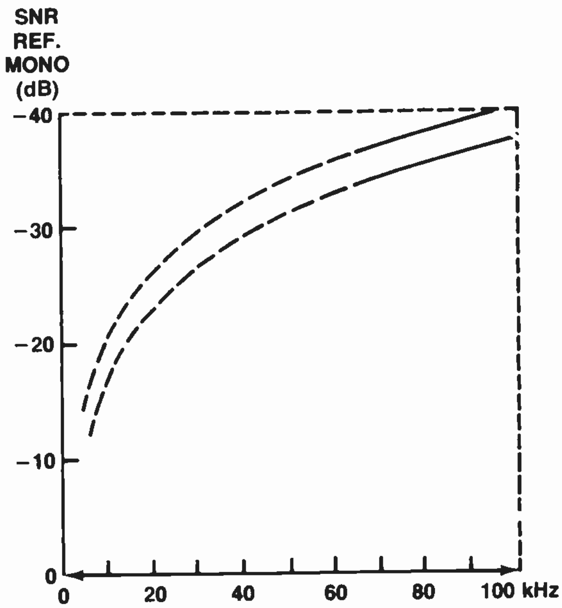


Figure 4.4-15. SNR of FM/SCA audio (5 kHz bandwidth) versus subcarrier frequency, compared to monophonic reception, referred to 5 kHz and 7 kHz deviation.

After frequency demodulation, the SNR power ratio will be:[15]

$$\mathrm{SNR} = \frac{3i^2}{\mathrm{CNR}} = \frac{3\,(1)^2}{\mathrm{CNR}} \qquad [14]$$

where CNR is the carrier to noise power ratio in the bandwidth occupied by the modulated signal. The noise power bandwidth B_n is difficult to predict for two reasons. First, the (usually) inexpensive SCA predetection filter has sloping skirt selectivity, thus departing from the ideal rectangular shape assumed in the calculation. Second, the actual bandwidth may vary, depending on the amount of audio distortion to be tolerated. The extremes of choice may be:

$$i \le 1,\ B_n = 2f_m = 10\ \text{kHz} \qquad [15]$$

$$i > 1,\ B_n = 2(f_d + f_m) = 20\ \text{kHz} \qquad [16]$$

A noise power bandwidth of 15 kHz is a compromise value that happens to represent the measured performance in some common FM SCA receivers.

In single sideband systems, one is simply frequency shifting the modulation passband to a particular frequency in the composite baseband. There is no inherent noise reduction effect; the recovered audio S/N ratio is approximately equivalent to the noise contained within the subchannel bandwidth.While the FM subcarrier system improves S/N, its greater bandwidth admits about three times the noise power. At a modulation index of one, FM improvement is approximately 10 log (1) or 0 dB for an ideal detector. The use of pre-emphasis and de-emphasis improves the FM system S/N without significantly affecting the subcarrier's occupied bandwidth. The use of 150 μsec de-emphasis, common in FM subcarrier demodulators, reduces RMS noise in a 5 kHz channel by 8.8 dB. The practical improvement in S/N ratio measured by the author is more typically 4 dB. This is because the extreme high frequency boost is not a good fit to the program material.[16]

Since single sideband modulation is spectrally efficient, even when compared to narrow band FM, it is practical to divide the 53–99 kHz range into a number of independent subchannels. In the one system, five subchannels each having a 5 kHz baseband are operated at 57, 66.5, 76, 85.5 and 95 kHz. These are shown in Figure 4.4-11.

Twenty percent peak injection is the maximum permitted by the FCC Rules when stereo is transmitted, therefore, each subchannel would receive 4% injection (20/5 = 4). It is reasonable to assume that slightly higher injection could be permitted for each subchannel due to the random interleaving of the total subchannel injection. If 5% were employed, then, the S/N would be lower than that for a channel using ten percent peak injection.

$$\text{change in SNR} = 20\log(10/5) = -6\ \text{dB} \qquad [17]$$

Assuming an approximated improvement of 4 dB for a standard FM subchannel, and a reduction of 6 dB for one of five single sideband subchannels, there is an estimated difference in S/N, assuming the same center frequency of both subchannels and flat noise density over this spectrum.

$$\mathrm{SNR_{FM}} - \mathrm{SNR_{SSD}} = 4 - (-6) \approx 10\ \text{dB} \qquad [18]$$

The power density from an AM demodulator remains uniform across the baseband (unlike FM demodulators, in which noise density rises in proportion to frequency). Therefore, pre-emphasis/de-emphasis is not as necessary to combat noise. The use of a complementary noise reduction system (compandor) can significantly improve the S/N ratio of either system.

Using the previous estimates, the performance of the (uncompandored) single sideband and (de-emphasized) FM subchannels is compared to main channel and stereo in Figure 4.4-16. The coverage radii are based on the FCC F(50,50) field strength predictions for an FM station having 10 kW ERP, 1000 ft height above average terrain (HAAT) and a dipole receiving antenna at 30 ft above ground level into a receiver having a 10 dB noise figure.

SCA-TO-STEREO INTERFERENCE CONSIDERATIONS

Introduction of new signals into the modulating baseband of an FM station requires attention to their possible interactions with existing baseband signals, as well as its chosen performance characteristics.

FM stereo receivers having stereo decoders with diode switching bridges driven by a high level 38 kHz sinusoid (usually identified by discrete component con-

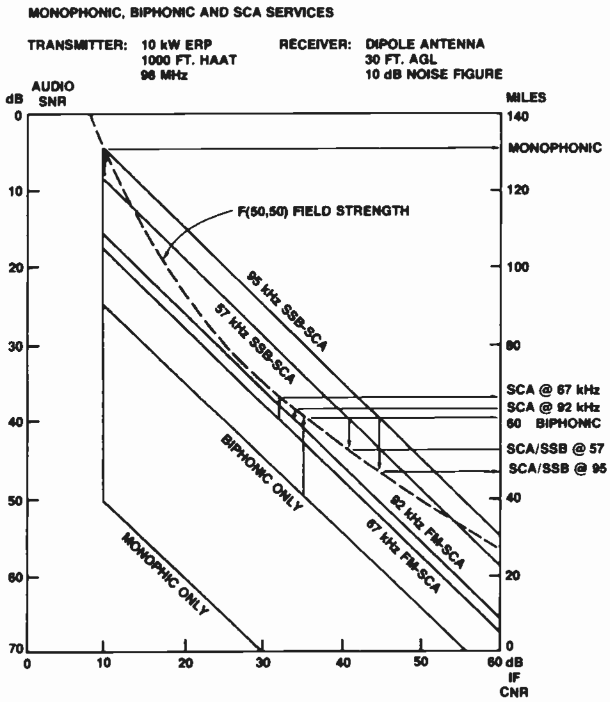


Figure 4.4-16. Predicted serviced distances for 50 dB monophonic and biphonic SNR, and 40 dB SCA SNR. To use, find intersection of desired SNR with service (diagonal) line; next move up (or down) to F(50,50) dashed line; then read across to mileage scale.

struction and tuned transformer stages) were responsible for the original birdie or beat note interference in early SCA/stereo broadcast systems. For historical perspective, the sources of this interference were:

- nonlinearity in the diode bridge due to the conduction characteristics of the diodes and the sinusoidal drive signal
- imbalance in the center tapped transformer driving the diode bridge and second harmonics of the 38 kHz sinusoid at this point

The result caused a second order intermodulation (mixing) of twice the 38 kHz switching frequency minus the instantaneous frequency of the SCA subcarrier as:

$$9\text{ kHz} = [38\text{ kHz} \times 2] - 67\text{ kHz} \qquad [19]$$

The product is not 10 kHz as was often believed. Today, the small amount of beat note interference that results from a 67 kHz subcarrier is 10 kHz, but due to an entirely different mechanism.

Over the past decade these receivers have been entirely replaced by sets with integrated circuit decoders, which have virtually no internally generated beat note problem. As a result, the interference to stereo service from SCA operation has dropped significantly.

Contemporary receivers create very little beat note interference from SCA operation because of the type of stereo decoding done within the integrated circuit. The fact that the circuit contains a PLL has little to do with this improvement. Modern decoders derive their 38 kHz signal by a digital technique, which is well suited to circuit integration. This results in a 38 kHz square wave of fast rise time and balanced duty cycle (equal time in high and low states). The output circuit is usually a differential transistor pair configured as a balanced demodulator. When driven by the 38 kHz square wave (which has virtually no even order harmonics), very little mixing is possible which can result in the product shown in the above equation.

If SCA beat notes created within stereo receivers was the only source of interference, there would be virtually no problem with SCA operation. However, experience shows that a minor source of interference remains when certain SCA frequencies are used. The source of this interference is *external* to the receiver: multipath distortion of the radio frequency signal itself.

Large obstructions such as mountains, hills and tall buildings can reflect VHF broadcast waves well enough to create the simultaneous reception of a direct and one or more reflected path signals. Upon demodulation the baseband signals will include a combination of linear, second and third order distortion products, as shown in Figure 4.4-17. Higher order distortions are negligible when the distortion is small. Hence, these products take the form:

2A, 2B, 2C, etc. (where A, B, and C are the fundamentals)

A + B, A − B, A + C, A − C, B + C, B − C, . . . etc.

2A + B, 2A − B, 2B + A, 2B − A, . . . etc.

Second order distortion (frequency sum and difference) is the most troublesome, because new product frequencies are created which may fall within the main and stereophonic channels where they can become audible. This is illustrated in Figure 4.4-18, which shows the baseband of an actual FM station. The only modulation present in the transmitter is the stereo pilot and unmodulated SCA subcarriers at 67 kHz and 92 kHz, plus a small amount of noise from the audio chain which can be seen from zero to 15 kHz and centered around 38 kHz in the $L-R$ subchannel.

Several second order products are visible in the spectrum graph. This distortion takes the form:

$$f_{SCA1} \pm f_{SCA2} = f_{IM} \text{ product, or} \qquad [20]$$
$$f_{SCA} \pm \text{pilot} = f_{IM} \text{ product}$$

In the case of 67 and 92 kHz subcarriers, significant products are:

92 − 67 = 25 kHz
67 − 19 = 48 kHz
92 − 19 = 73 kHz
67 + 19 = 86 kHz
92 + 19 = 111 kHz

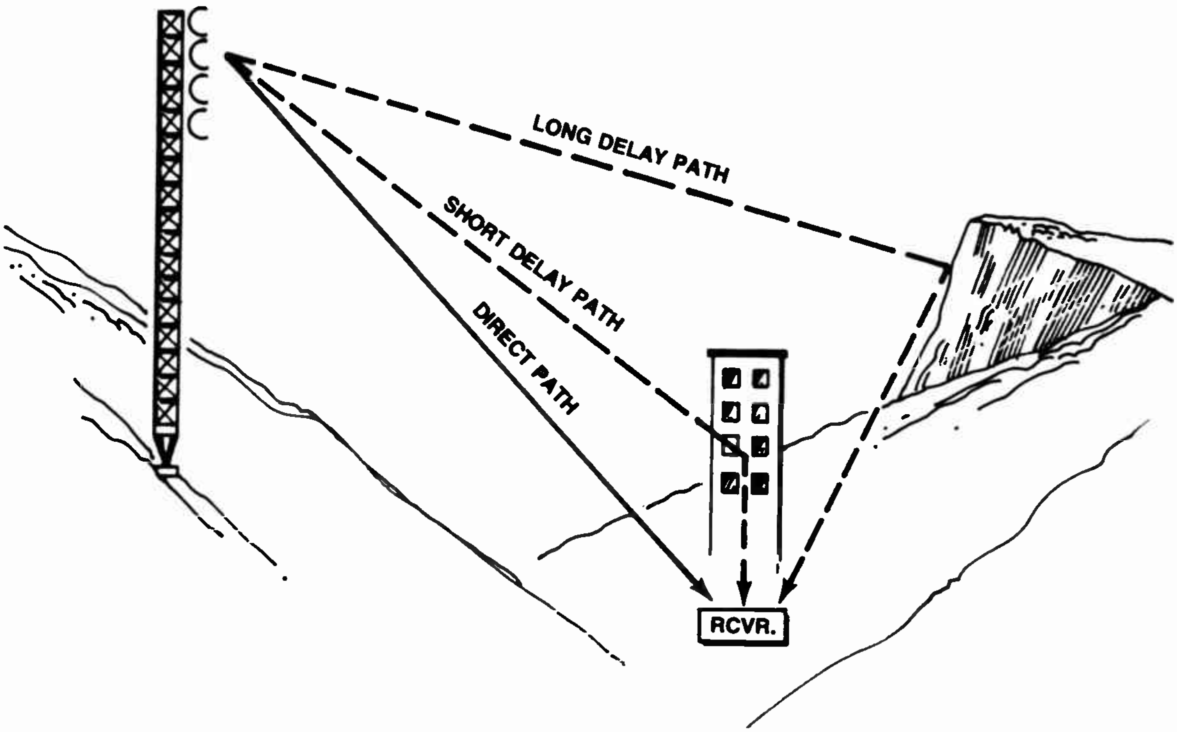


Figure 4.4-17. Examples of reflected signals arriving at an FM receiver, along with a direct path RF signal. Distortion of the demodulated FM signal is related to both the direct/reflected signal ratio and delay time of the reflected signal; distortion increases as the signal ratio approaches 1:1 and as the secondary path delay time rises.

While these IM products are not directly audible, some may be once they are demodulated by a stereo decoder:

$$25 - 38 = 13 \text{ kHz}$$
$$48 - 38 = 10 \text{ kHz}$$
$$111 - 114 = 3 \text{ kHz}$$

(The last product could be demodulated by IC stereo decoders using a 38 kHz square wave switching signal. However, sensitivity to a 111 kHz product is rare.)

Some other general findings about multipath effects are:[18]

- Multipath distortion is almost inversely proportional to the DU (desired-to-undesired path signal ratio) if the ratio is greater than about 10 dB
- Baseband signal distortion increases almost proportionally to the delay time up to about 10 μsec, then increases with delay time at a slower rate
- A high DU ratio required to suppress the beat note interference to stereo program reception from an SCA subcarrier (about 20 dB at a delay time of around 5 μsec and about 30 dB at around 20 μsec).

It should be emphasized that the perceptibility of this form of beat note interference depends on the loudness and consistency of main channel programming (program audio in most broadcast FM stations is capable of masking the beat note). Generally, only stations which do light amounts of audio processing and broadcast programming with wide dynamic range are even aware of an occasional case of interference.

The *type* of SCA modulation or the subcarrier frequency may significantly affect the perceptibility of any beat note. For example, properly encoded high speed data subcarriers using FSK or PSK spread the carrier energy in such a way that it sounds like faint, band limited noise which is most easily masked. Fur-

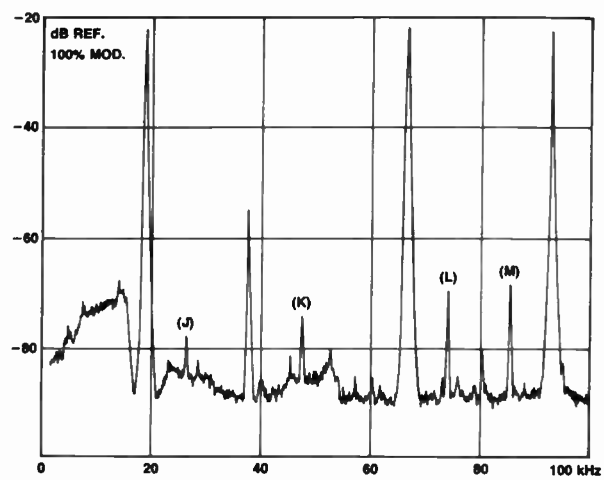


Figure 4.4-18. Composite baseband of an FM transmitter system with pilot, 67 kHz, and 92 kHz SCAs at 9% injection. IM products are identified by letters and are described in the text.

thermore, higher subcarrier frequencies will shift the beat-note image out of the audible range. Experience has shown that 92 to 95 kHz subcarriers have little problem in this regard.

Crosstalk From Main Channel Program Into SCA Subchannels

Crosstalk remains a problem for audio SCA services, despite the development of low distortion receivers for SCA use. The cause is the same as SCA-to-stereo beat notes: multipath distortion.

In the reference case illustrated in Figure 4.4-19, the baseband signal consists of one low frequency sinewave of frequency F (1 kHz), which modulates the main carrier by a deviation D (95% or 71.25 kHz for broadcast FM) and a sinewave subcarrier at a variable frequency f Hz which modulates the main carrier with a very small deviation d, having a modulation index less than 0.3.

Second order distortion will cause sidebands at frequencies of $f-F$ and $f+F$ with an output injection $d_2(f-F)$ and $d_2(f+F)$. The sum of the sideband amplitudes is:

$$d_2(f) = d_2(f - F) + d_2(f + F) \qquad [21]$$

The distortion is defined as the ratio of the sum of the sidebands to the amplitude of the subcarrier d_1 is:

$$\delta = d_2(f)/d_1(f) \qquad [22]$$

This distortion is easily measured as a function of the subcarrier frequency f with a spectrum analyzer. If the sidebands are equal they may in the extreme either amplitude modulate or phase modulate the subcarrier. In general there is a combination of amplitude and phase modulation.

Third order distortion is similarly defined by the sidebands created at frequencies $f-2F$ and $f+2F$.

While FM SCA detectors usually include amplitude limiting, they are intended to convert any angular modulation, whether program audio or not, into an output signal. Thus any phase non-linearity in the RF system or multipath will generate second and higher order sidebands around the subcarrier, causing main-to-SCA crosstalk at audio modulating frequencies.

Figure 4.4-20 shows the relationship between main-to-SCA crosstalk, subcarrier frequency, and maximum

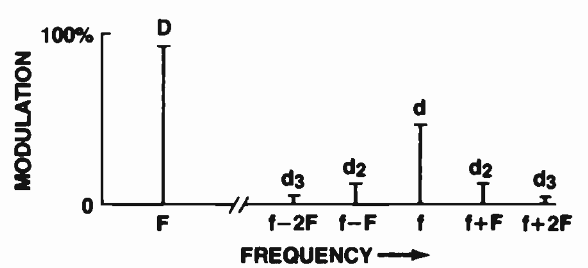


Figure 4.4-19. Spectrum of a subcarrier "f" shown with second and third order distortion products resulting from low frequency modulation "f".

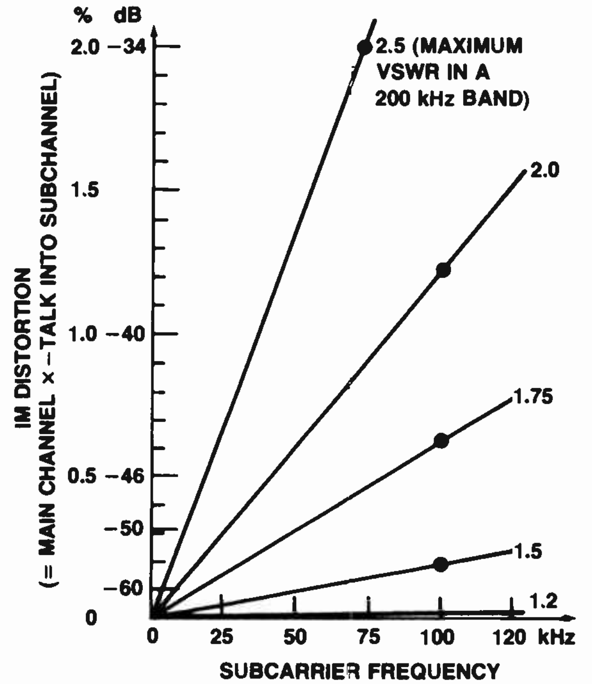


Figure 4.4-20. Distortion due to limited antenna bandwidth vs. frequency of subcarrier with max VSWR in 200 kHz band as parameter.

antenna VSWR in a 200 kHz band. It is evident that higher frequency subcarriers require somewhat larger, but not very much larger bandwidth. It has also been determined that antenna matching must be improved as the transmission line becomes longer. This is especially important for higher frequency subcarriers since the phase error is compounded rapidly with an increase in subcarrier frequency. Distortion is proportional to the subcarrier frequency. Thus, for equal distortion, doubling the subcarrier frequency requires halving the reflection coefficient.[19]

REFERENCES

1. Armstrong, Edwin H., "A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation," *Proceedings of the I.R.E.*, May 1936, Vol. 24, No. 5, pp. 689–740.
2. Electronic Industries Association, *Report on the National Quadraphonic Radio Committee to the Federal Communications Commission*, Vol. II, Chapter 1, pp. 21–29.
3. Federal Communications Commission, "Rules and Regulations," 47 C.F.R. Section 73.310.
4. National Association of Broadcasters/Westinghouse Broadcasting & Cable, Inc./National Public Radio, *Increased FM Deviation, Additional Subcarriers and FM Broadcasting: A Technical Report*, Aug. 30, 183, pp. 5–6.
5. Takahasi, S., Iida, H., "Application of Walsh

Functions to an FM Stereo Demodulator," J. Audio Eng. Soc., Vol. 33, No. 9, September 1985.
6. FCC BC Docket No. 82-536, *First Report and Order* (Apr. 7, 1983), *Second Report and Order* (Mar. 2, 1984).
7. Kean, John C., "Laboratory and Field Tests of Several FM/SCA Frequencies," *National Public Radio Engineering Report,* Oct. 15, 1981, pp. 13–14.
8. Moseley Associates, *Technical Manual for Model SCG-8 FM Subcarrier Generator.*
9. Small, Eric and Stoll, Eric., "Data SCA: Some Real World Experience," Modulation Science Pamphlet, Aug. 1984.
10. Sams, Howard W. Inc., *Reference Data for Radio Engineers, 6th Edition,* 1975, pp. 23-9–23-12.
11. Ibid., NQRC, Vol. II-1, Chapter 1, pp. 31–40.
12. Ibid., NQRC, Supplement to Volume II, Chapter 1, Section 1.4, p. xviii.
13. Ibid., pg. 36.
14. Ibid., Kean, pg. 24.
15. Ibid., Sams, pp. 23-4–23-12.
16. Ibid., Kean, pg. 24.
17. Ibid., NQRC, papers of James Gibson, RCA, Chapter III, pg 3/7.
18. Mitsuo O'Hara, "Distortion and Crosstalk Caused by Multipath Propagation in Frequency-Modulated Sound Broadcasting," *IEEE Transactions on Broadcasting,* Vol. BC-26, No. 3, Sept. 1980, pp. 76–81.
19. Ibid., NQRC/Gibson, pg. 3/11.

4.5
STL SYSTEMS

JERRY C. WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

One of the major concerns in the design and operation of a radio broadcasting facility is the means by which the program audio from the studio is conveyed to the transmitter site. As illustrated in Figure 4.5-1, this link represents an important element in the overall reliability of the transmission chain. Furthermore, as digital technology continues to move into daily radio station operation, the *studio-to-transmitter link* (STL) must become as transparent as possible. An inferior link will impose an unacceptable limit on overall audio quality. The requirements for reliability and transparent program relay have led to the development of new STL systems based on digital technology.

Changes in FCC broadcast ownership rules and the popularity of *local marketing agreements* (LMAs) have reshaped radio broadcasting. The need for high quality audio programming is one outgrowth of new competition and new alliances and STL systems are an important component of these audio improvement efforts. Furthermore, increasing numbers of stations are using *intercity relay* (ICR) facilities to share programming. Unfortunately, in many areas of the United States, the demand for 950 MHz STL channels has far outstripped the available supply. The Federal Communication Commission's (FCC) Part 74 bandwidth allocations for STL systems, therefore, necessitate highly efficient designs to meet the needs of radio broadcasters today.

Market demand for STL systems is not limited to North America. The commercialization of radio broadcasting in Europe and elsewhere has led to increased use of radio links to relay programming from one studio to the next, and from the studio to the transmitter site. In some areas, repeated use of the same frequency places stringent demands on system design.

Radio vs. Landline Links

The arguments over which approach—radio STL or landline (telco)—is the better way to convey program audio from the studio to the transmitter is as old as radio broadcasting itself. Now, however, a new element has been added to the equation that makes the landline option considerably more attractive. That new element is digital technology.

For years, broadcasters were limited in their choice of STL equipment to two main options: 950 MHz radio and equalized analog telephone lines. Today, a third approach, T1 digital transmission, is rapidly becoming the system of choice for many radio stations.[1] The decision to use T1 may be based on several factors, including:

- Necessity. The station has no line-of-sight to the transmitter, or suitable frequencies are unavailable

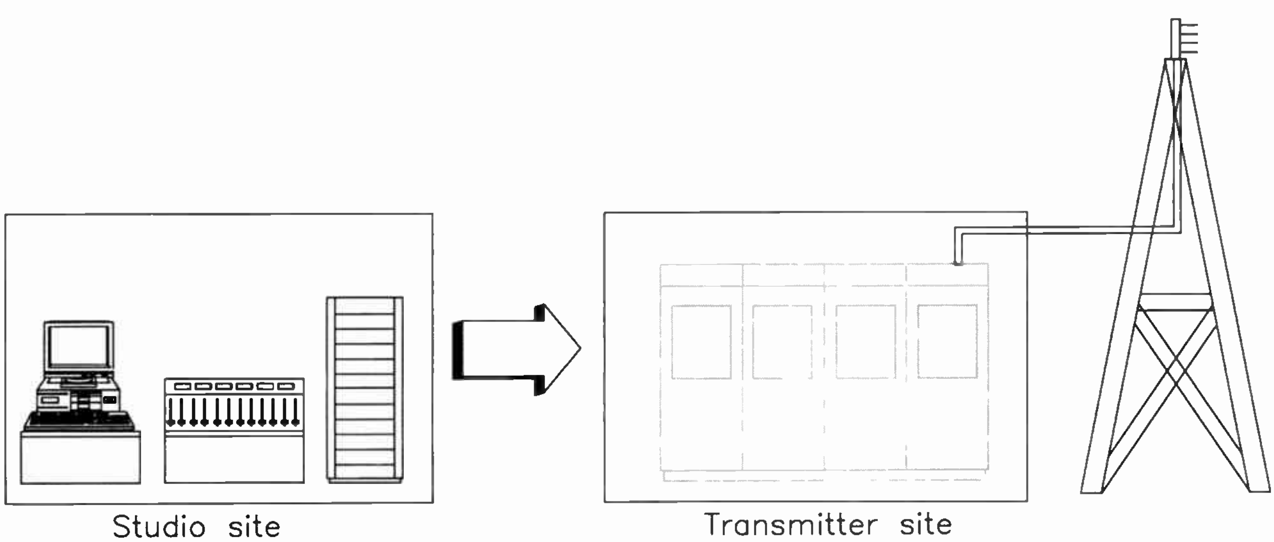


Figure 4.5-1. The basic function of a studio-to-transmitter link.[2]

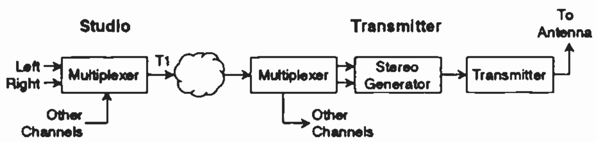


Figure 4.5-2. Discrete channel T1 digital audio STL system.[1]

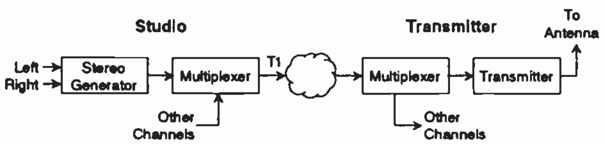


Figure 4.5-3. Composite T1 digital audio STL system.[1]

- Sound quality. A digital STL can sound better than even the best analog systems
- Cost. A single leased T1 line can cost less than multiple leased analog lines.

Whatever the reasons, a broadcaster who decides to use T1 transmission must then choose what type of system to implement. Digital T1 STL systems can be designed either to transmit discrete left and right channel stereo, as shown in Figure 4.5-2, or to transmit a composite stereo signal as illustrated in Figure 4.5-3.

One of the advantages to a broadcaster in using a T1 digital STL is the ability to multiplex several signals together. Subsidiary Communications Authority (SCA), second audio program (SAP) channels, transmitter remote control signals, data and voice can all be combined with the broadcast audio signal for transmission on a single T1 circuit. The duplex nature of T1 permits use of the same system for both STL and *transmitter-to-studio* (TSL) functions as well. Table 4.5-1 compares T1 bandwidth usage for various circuit implementations.

A detailed discussion of telco based STL systems is beyond the scope of this chapter. Interested readers are referred to Chapter 3.11, *Common Carrier Audio Program Services*, for a discussion of this subject.

STL SYSTEM CONFIGURATION

The frequency range used by aural STL systems in the United States is 944.5–951.5 MHz. Frequencies ranging from 200–940 MHz are used in other parts of the world. STL systems are also operated above these traditional bands, however, such links often face difficulties associated with increased susceptibility to signal loss as a result of precipitation, foliage and fog, which limit usage to short path lengths. Frequency modulation is used for analog STL systems.

The vast majority of radio STL systems in operation today carry a composite stereo baseband signal from the studio to the transmitter. An alternative approach, and one that may be used for monaural AM stations, involves the use of one or two narrowband STL systems. The two approaches are compared in Figure 4.5-4.

The monaural STL typically has an audio bandwidth of up to 15 kHz for program signals and usually can accommodate a single FM subcarrier at approximately 39 kHz as shown in Figure 4.5-5(a). For stereo or dual mono installations, a common antenna can be used at the transmitter and at the receiver. At the transmit site, a combiner/isolator adds the signals and prevents the generation of intermodulation products. At the receive site, a low-noise antenna splitter feeds the individual receivers.

The composite STL has sufficient bandwidth (220 kHz) to accommodate the output of a stereo FM generator and usually several separate subcarriers. A low pass filter generally is used at the output of the receiver to separate the signals to be broadcast over-the-air from the non broadcast (or closed-circuit) signals. A typical configuration is illustrated in Figure 4.5-5(b). The composite STL provides superior stereo performance compared to dual monaural radio links in several respects, including:

- Elimination of interchannel phase and amplitude errors that can arise in a dual channel system
- Elimination of audio headroom considerations because the STL input signal has already been passed through the station's audio processing system and the stereo generator, which are located at the studio.

For FM applications, the composite signal from the

Table 4.5-1
Digital Audio Sampling Size vs. T1 Bandwidth Usage.[1]

System Type	Samples/s (for 15 kHz stereo)	Sample Size	Digital Transmission Rate	Number of TI Time Slots Used for Broadcast Audio	Number of T1 Time Slots Available for Other Channels
Discrete	32,000 × 2	16 bit	1.152 Mbits/s	18	6
Composite	112,000	16 bit	1.792 Mbits/s	Not usable	—
		15 bit	1.680 Mbits/s	Not usable	—
		14 bit	1.568 Mbits/s	Not usable	—
		13 bit	1.456 Mbits/s	23	1
		12 bit	1.344 Mbits/s	21	3

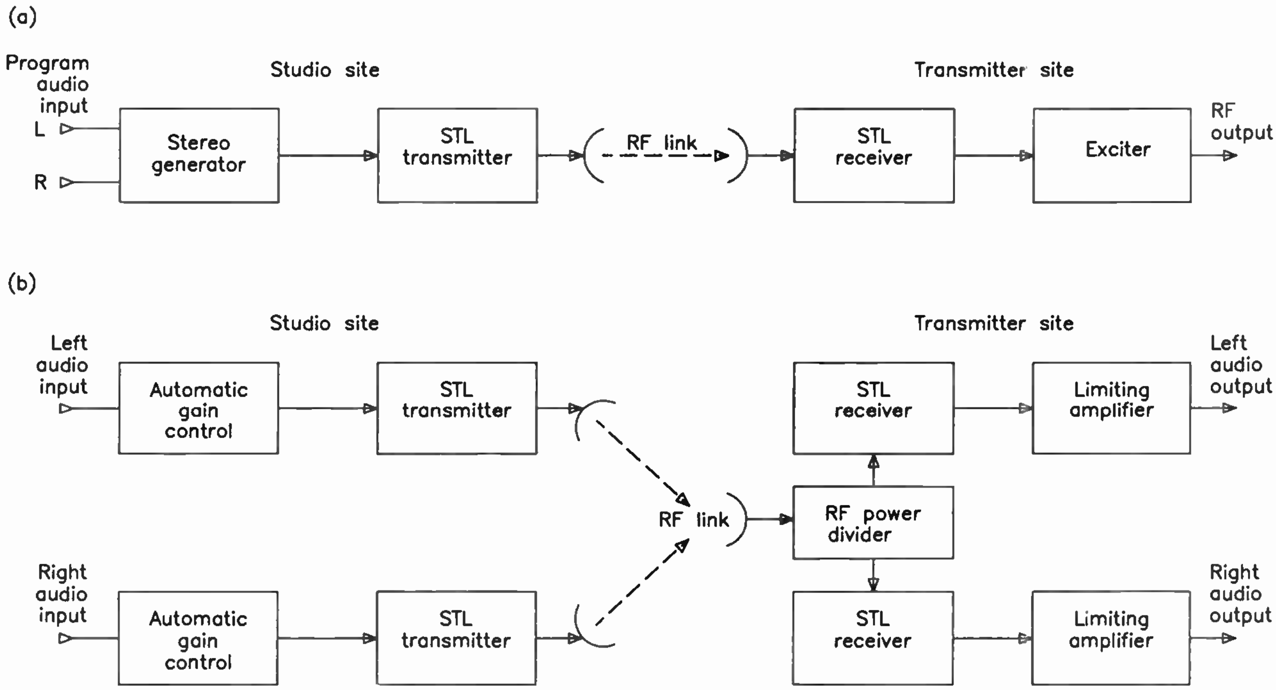


Figure 4.5-4. Comparison of STL systems: (*a*) composite transmitter-receiver system, (*b*) dual monaural transmitter-receiver system.[2]

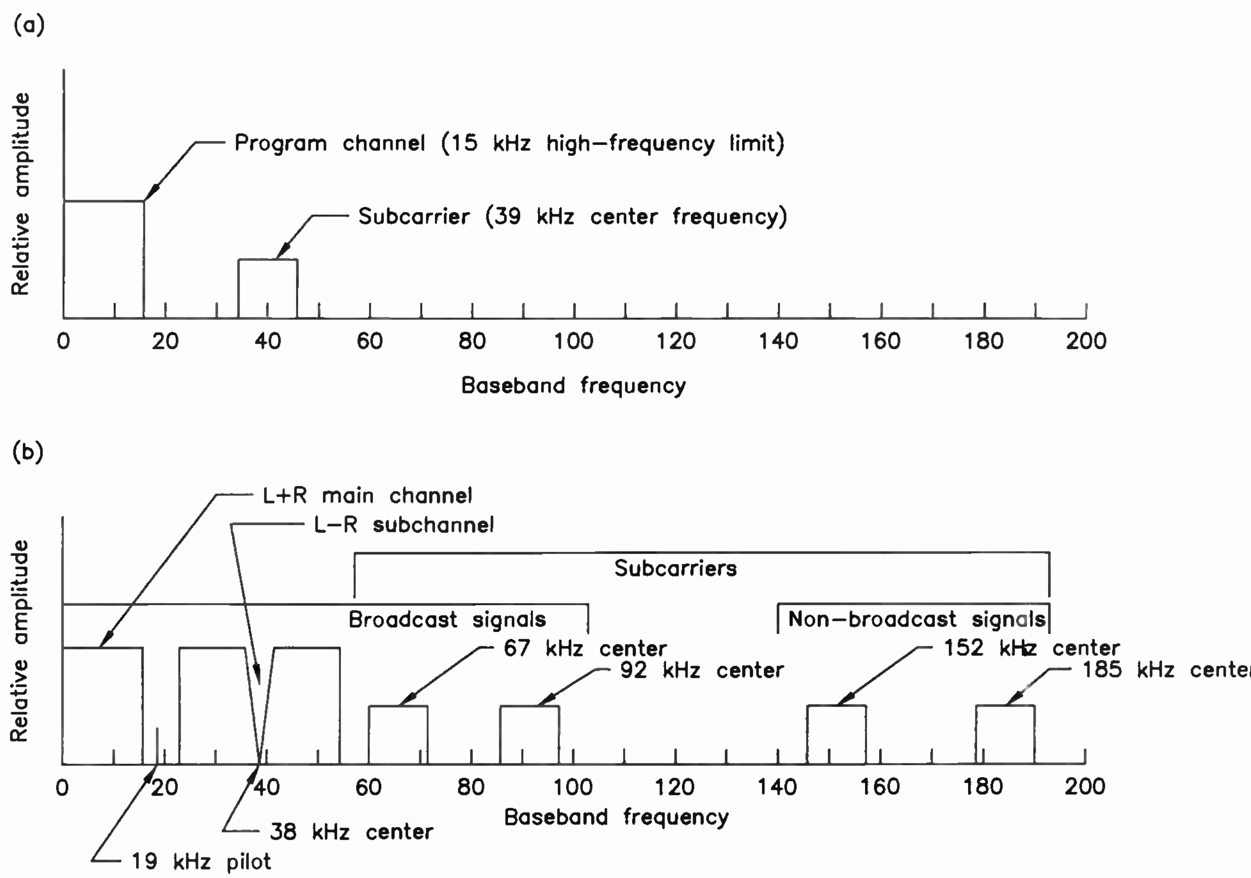


Figure 4.5-5. Baseband spectrum of STL systems: (*a*) monaural, (*b*) composite.[2]

STL receiver is fed directly into the FM transmitter exciter. An alternate method involves the use of a *Reciter*, which functions as an STL receiver and FM exciter but with an IF interface rather than a baseband interface.[2] The elimination of one stage of demodulation and remodulation reduces noise and distortion in the transmitted FM signals.

Because AM stereo exciters require discrete right and left channel audio inputs, a complementary stereo decoder is used to recover the left and right channel audio if a composite STL (fed by a stereo generator) is used.

Digital STL vs. Analog STL Link

Digital signal encoding of audio frequency information provides a number of advantages over conventional analog systems. The principal strengths of digital coding for transmission and storage vs. analog methods include the following:

- Greater immunity to noise and interference in the transmission path. Differences of 50 dB SNR vs. 76 dB are common
- Elimination of transmission-path-dependent distortion mechanisms, such as harmonic distortion, intermodulation distortion, and crosstalk
- Efficient use of baseband and radio frequency (RF) spectrum
- Efficient and predictable regeneration of the digital signal
- Easy and effective encryption for security and coding purposes
- Uniformity in transmitting audio and data signals.

The last item merits additional consideration. Intense development has been underway for a number of years to develop efficient and cost-effective digital coding and compression systems for audio frequency signals. The results of this work can be seen in everything from personal computer soundboards to recordable CD type minidisks. This technology, obviously, has many spinoff effects that can be used to advantage by radio broadcasters. Advanced digital devices and algorithms are now being adapted for use in a wide variety of broadcast products, most notably studio storage and mixing systems and STL systems.

The use of digital coding and modulation in an STL system provides the user a number of benefits in terms of performance and path reliability. By its nature, digital coding offers a more robust link. As long as the received signal strength and carrier-to-noise exceed a defined minimum value, the *bit error rate* remains relatively constant, and the decoded baseband signal is an exact duplicate of the input signal at the studio (see Figure 4.5-6). With adequate error correction, signals can be received error-free. With a digital STL, multiple links are practical because signal degradations are not additive, as in an analog system.

It is fair to point out that digitization of the input audio signal always brings with it a measure of degra-

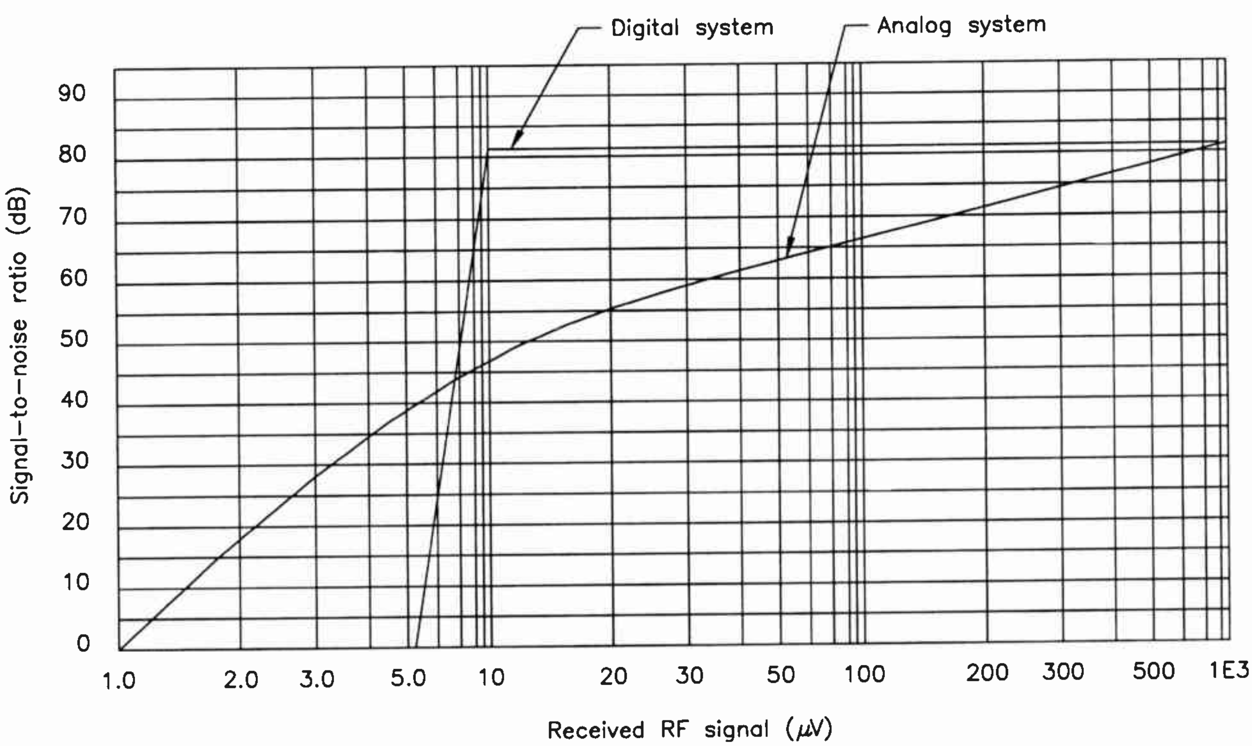


Figure 4.5-6. The benefits of digital vs. analog STL systems in terms of S/N and received RF level.[2]

dation in the form of quantization errors. However, with the high sampling rates typically used for professional audio applications, such degradation is minor and almost always completely inaudible.

The process of quantization is illustrated in Figure 4.5-7. The sampling rate and quality of the sampling circuit determine, in large part, the overall quality of the digital system. A properly operating transmission channel can be assumed to provide error-free throughput. This being the case, the digital signal can be regenerated at the receiving point as an exact duplicate of the input waveform, quantization errors excepted. Figure 4.5-8 shows a general representation of a digital communications channel. In the case of a radio link, such as an STL, the transmission medium is analog in nature (FM). The circuits used to excite the modulator, however, are essentially identical to those used for an all-digital link, such as fiber optic cable.

The functions of the encoder and decoder, shown in Figure 4.5-8(a), usually are formed into a single device, or set of devices (a chip set), known as a *codec* (*coding and decoding* device). At the transmission end, the codec provides the necessary filtering to band-limit the analog signal to avoid aliasing, thereby preventing *analog-to-digital* (A/D) conversion errors. At the receiver, the codec performs the reciprocal *digital-to-analog* (D/A) conversion and interpolates (smoothes) the resulting analog waveform.

Digital STL Performance

The benefits of a digital STL for radio broadcasters can best be appreciated by comparing the performance of a digital system and an analog system. A digital STL typically permits broadcasters to extend the fade margin of an existing analog link by 20 dB or more. Furthermore, audio *signal-to-noise* (S/N) improvements of at least 10 dB can be expected for a given RF signal strength. Alternatively, for the same S/N, the maximum possible path distance of a given composite STL transmitter and receiver can be extended. These features could, in some cases, mean the difference between a one-hop system and a two-hop system.

The spectrum-efficiency of a digital STL is of great importance today in highly congested markets. The system may, for example, be capable of relaying four program channels and two voice-grade channels. The use of digital coding also makes the signals more tolerant of co-channel interference than a comparable analog STL.

Coding System

Several approaches may be used to digitize or encode the input audio signals. The complexity of the method used is a function of the availability of processing power and encoder memory and determines the resulting delay incurred due to the encoding/decoding process. For a real-time function such as an STL, significant encoding/decoding delay is unacceptable. *Pulse code modulation* (PCM) is a common scheme that meets the requirements for speed and accuracy. In the PCM process, the sampled analog values of the input waveform are coded into unique and discrete values. This quantization may be uniform, as illustrated in Figure 4.5-7, or nonuniform. With nonuniform quantization, compression at the coder and subsequent expansion at the decoder is performed. By using larger quantization steps for high energy signals and smaller steps for low energy signals, efficient use is made of the data bits, while maintaining a specified signal-to-quantization noise level. This process is known as *companding* (compression and expansion).

PCM encoding, in a simple real-time system, provides a high-speed string of discrete digital values that represent the input audio waveform. Each value is independent of all previous samples. No encoder memory is required. This approach, while simple and fast, is not particularly efficient insofar as the transmission channel is concerned. There are many redundancies in any given input signal. By eliminating the redundancies, and taking advantage of the *masking* effects of human hearing, greater transmission efficiency can be realized. Viewed from another perspective, for a given radio transmission bandwidth, more information can be transferred by using a compression system that

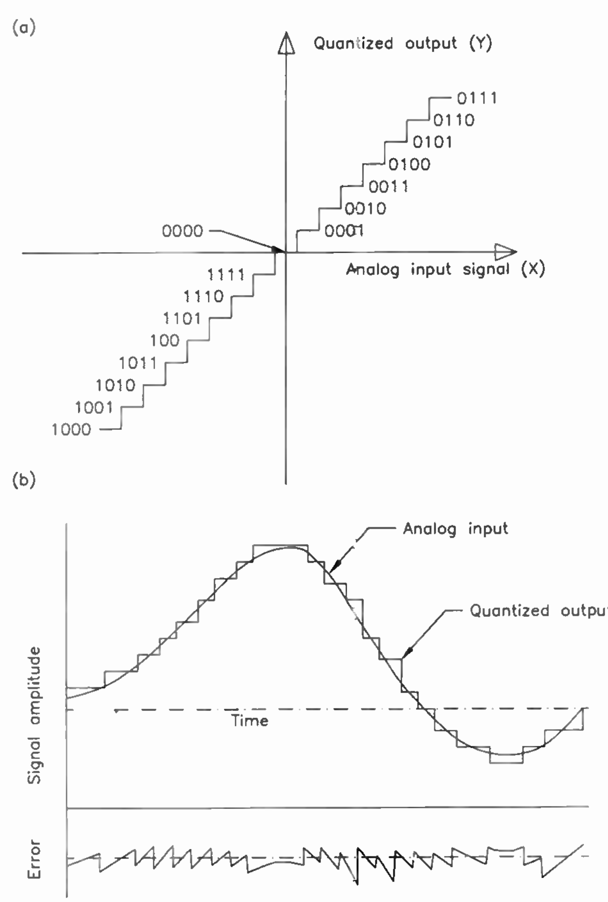


Figure 4.5-7. Quantization of an input signal: (*a*) quantization steps, (*b*) quantization error signal.[2]

removes nonessential data bits. (Audio compression systems are discussed in Chapter 3.7, *Digital Audio Compression Technologies.*)

Analog Composite STL Transmitter and Receiver Characteristics

As mentioned previously, the composite aural STL is the workhorse of the radio industry. A number of hardware variations can be found in transmitter and receiver systems, but all are designed to accomplish basically the same goal: relay of a baseband signal from the studio site to the transmitter. Figure 4.5-9 illustrates the basic STL configurations for an AM stereo (Figure 4.5-9(a)) and FM stereo (Figure 4.59(b)) application. A block diagram of a representative composite STL is given in Figure 4.5-10.

The RF carrier is generated by a *voltage-controlled oscillator* (VCO) that is phase-locked to a modulating VCO operating at an intermediate frequency of approximately 70 MHz. Direct composite FM modulation takes place at the IF VCO, which is also phase-locked to an oven-controlled crystal oscillator that serves as the master timebase for the transmitter. Because fre-

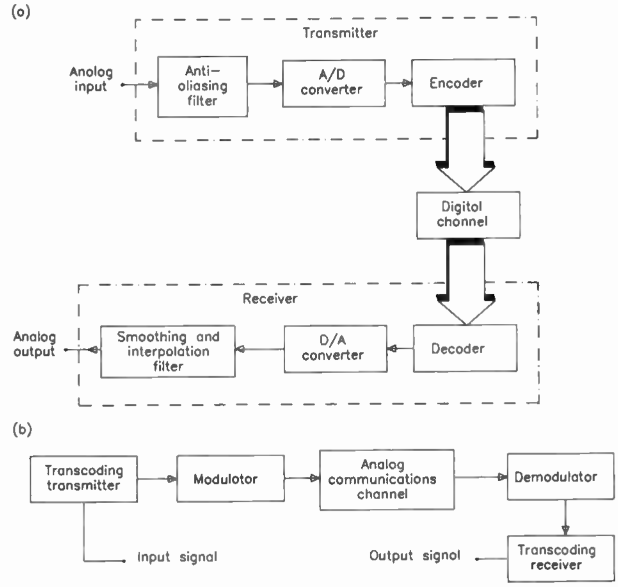


Figure 4.5-8. Digital transmission system: (*a*) coding/decoding functions, (*b*) overall communications link.[2]

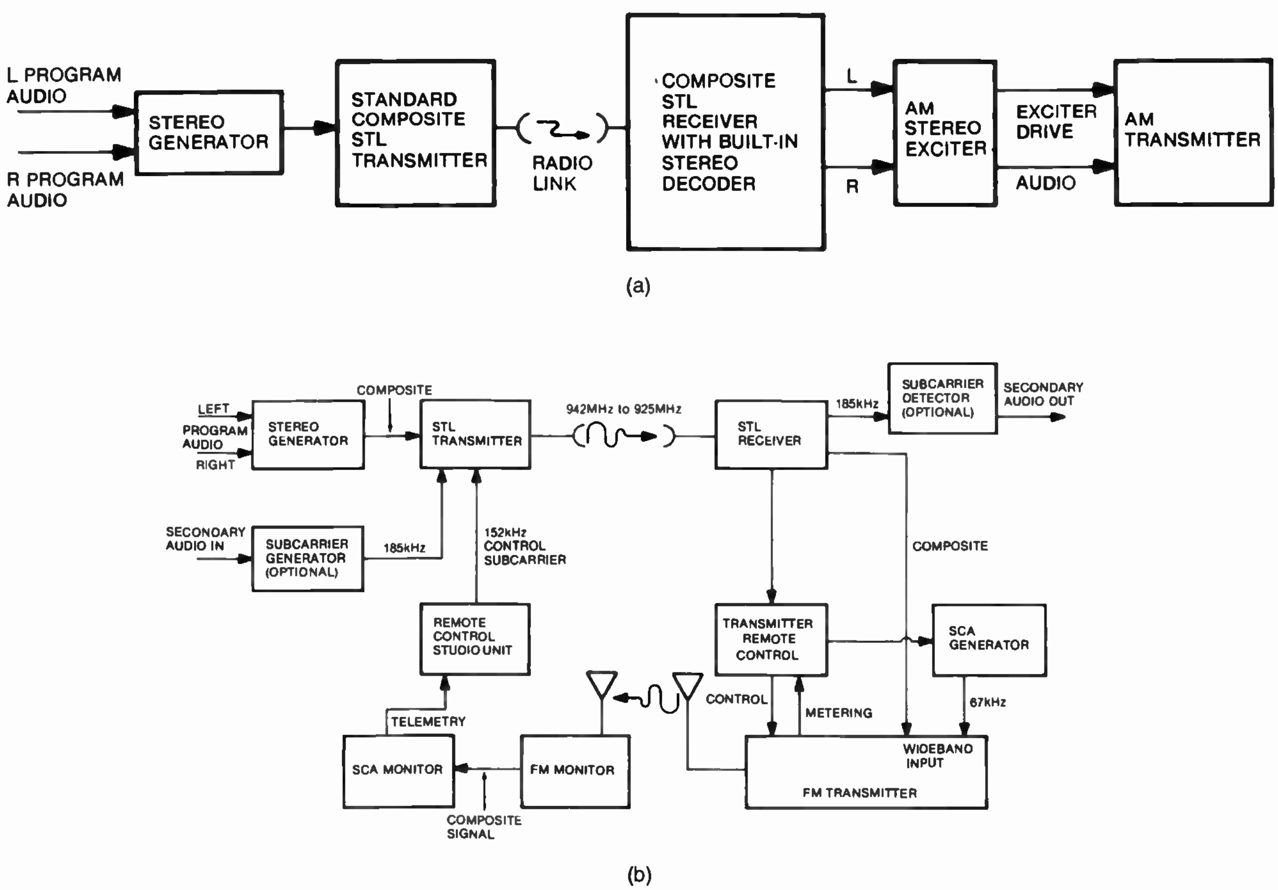


Figure 4.5-9. Analog composite STL links: (*a*) AM radio application, (*b*) FM radio application.[2]

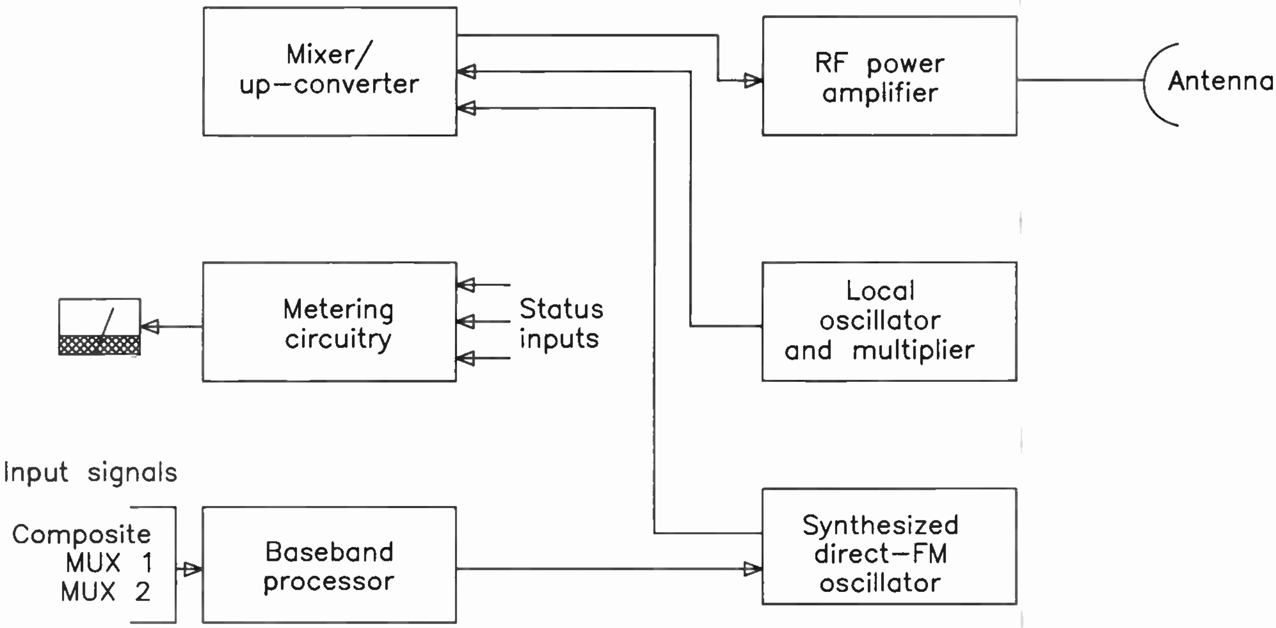


Figure 4.5-10. Block diagram of a composite STL transmitter.[2]

quency multiplication is not used, the RF carrier is an exact reproduction of the IF VCO. Therefore, a number of potential compromises resulting from frequency-multiplication and post-heterodyne filtering are eliminated, including:

- Degradation of S/N
- Generation of spurious signals
- Degraded stereo performance

IF modulation provides excellent overall stereo performance, as documented in Table 4.5-2.

In this design, the operating frequency of the transmitter is programmable by DIP switches in 12.5 kHz steps, so that frequency changes may be made in the field. The composite baseband signal accommodates the full stereo baseband, including two SCA and/or multiplex (MUX) channels. These features permit all processing equipment for analog operation to be located at the studio. An illustration of the composite baseband of such a system is shown in Figure 4.5-11.

Table 4.5-2
Specifications for a Typical Composite STL System.[2]

Parameter	Specification
Power output	6 to 8 W
Frequency stability	±0.0002%, 0 to 50°C
Spurious emissions	60 dB below maximum carrier power
Baseband frequency response	±0.1 dB or less, 30 Hz to 75 kHz
Stereo separation	Greater than 55 dB at 1 kHz
Total harmonic distortion	0.02%, 75 μs deemphasis
S/N	85 dB below ±75 kHz deviation, 75 μs deemphasis
Nonlinear crosstalk	50 dB or less
Subchannel-to-main crosstalk	60 dB or less

A block diagram of the companion composite STL receiver is shown in Figure 4.5-12. Like the transmitter, the receiver is user-programmable in 12.5 kHz

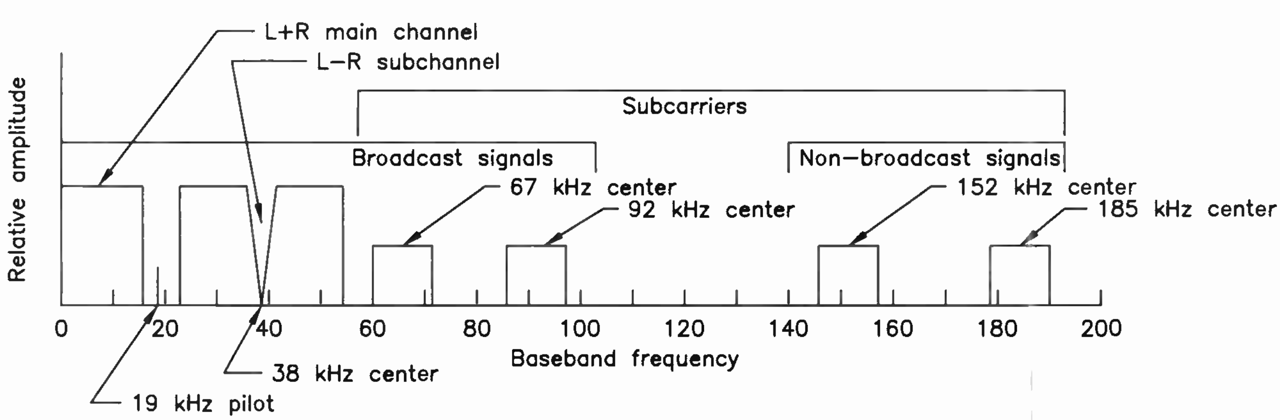


Figure 4.5-11. The baseband of a composite STL transmitter.[2]

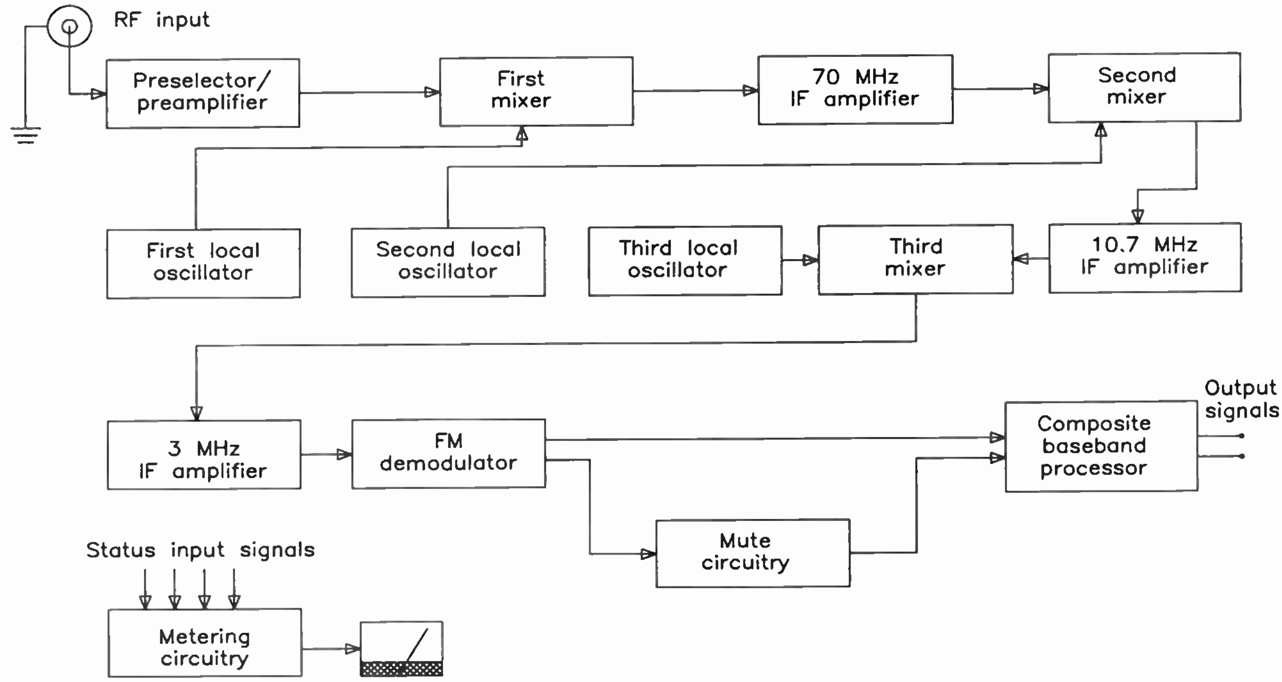


Figure 4.5-12. Block diagram of a composite STL receiver.[2]

steps typically through the use of internal dual inline package (DIP) switches. The front-end uses cascaded high-*Q* cavity filters and *surface acoustic wave* (SAW) IF filters to provide high selectivity and phase linearity. Triple conversion IF is used to feed a pulse-counting discriminator for linear baseband demodulation.

Components of an STL System

An STL installation is only as good as the hardware used to interconnect the link. All components, from the transmitter to the output connectors, must be carefully chosen and properly installed. A well-designed system will provide years of trouble-free service. A poorly designed system will cause problems on a regular basis.

Transmitter and Receiver Requirements

When choosing an STL transmitter, it will be necessary to select a unit that will deliver sufficient power to overcome the losses determined by path gain/loss calculations. Figure 4.5-13 provides a starting point for system planning. In cases where additional power is needed, a separate power amplifier may be used. For convenience in manipulating figures, the transmitter power output should be converted to gain in decibels above a 1 mW reference (dBm) (see Table 4.5-3).

When choosing an STL receiver, specifications should be carefully analyzed, particularly receiver sensitivity. This figure, necessary in STL path calculations, is usually specified as a signal level required for a specified S/N. This value should be converted to dBm (see Table 4.5-4). For example, a receiver may require 100 μV for 60 dB S/N. This is equivalent to −66.9 dBm. In receiver design, sensitivity, S/N, selectivity and the method of demodulation are deter-

Example: H_T = 60 feet. H_R = 500 feet. Radio path length = 41.5 miles

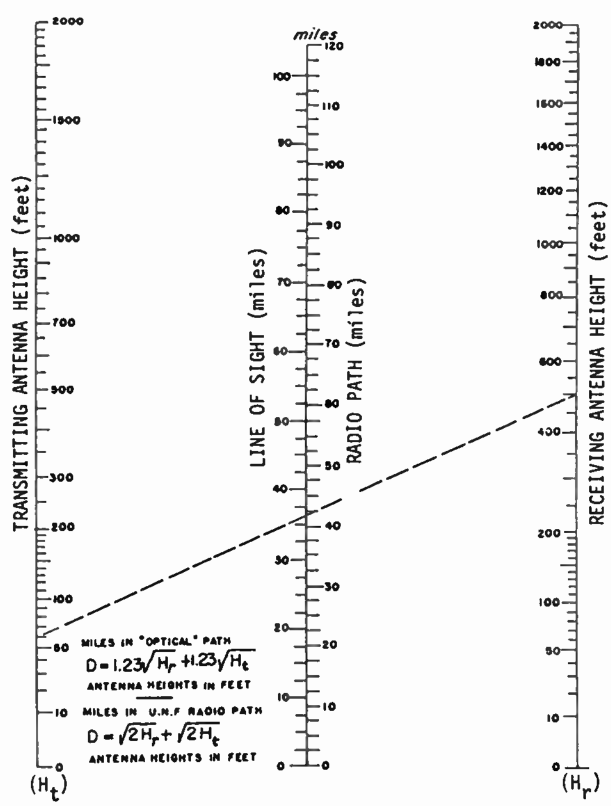


Figure 4.5-13. Nomogram to quickly determine if a path is viable.[2]

Table 4.5-3
Conversion of Watts to dBm (50 Ω impedance).

Power in watts	dBm	Power in watts	dBm
5.0	37.0	18.0	42.6
5.5	37.42	19.0	42.8
6.0	37.79	20.0	43.0
6.5	38.12	25.0	43.9
7.0	38.46	30.0	44.8
7.5	38.75	35.0	45.4
8.0	39.0	40.0	46.0
8.5	39.3	45.0	46.5
9.0	39.55	50.0	47.0
9.5	39.78	55.0	47.4
10.0	40.0	60.0	47.8
10.5	40.2	65.0	48.1
11.0	40.4	70.0	48.5
12.0	40.8	75.0	48.8
13.0	41.1	80.0	49.0
14.0	41.5	85.0	49.3
15.0	41.8	90.0	49.5
16.0	42.0	95.0	49.9
17.0	42.3	100.0	50

Table 4.5-4
Conversion of Microvolts to dBm (50 Ω impedance).

Microvolts	dBm	Microvolts	dBm
0.10	−127	40	−74.9
0.12	−125.25	45	−73.9
0.14	−124	50	−72.9
0.16	−122.9	60	−71.25
0.18	−121.9	70	−70
0.20	−120.9	80	−68.9
0.25	−119	90	−67.9
0.30	−117.25	100	−66.9
0.35	−116	120	−65.25
0.40	−115	140	−64
0.60	−111.25	160	−62.9
0.70	−110	180	−61.9
0.80	−108.9	200	−60.9
0.90	−107.9	250	−59
1.0	−106.9	300	−57.5
1.2	−105.25	350	−56
1.4	−104	400	−54.9
1.6	−102.9	450	−53.9
1.8	−101.9	500	−52.9
2.0	−100.9	600	−51.25
2.5	−99	700	−50
3.0	−97.5	800	−49
3.5	−96	900	−48
4.0	−95	1000	−46.9
4.5	−93.9	1200	−45.25
5.0	−92.9	1400	−44
6.0	−91.25	1600	−42.9
7.0	−90	1800	−41.9
8.0	−88.9	2000	−40.9
9.0	−87.9	2500	−39
10	−86.9	3000	−37.25
11	−86	3500	−36
12	−85.25	4000	−34.9
14	−84	4500	−33.9
16	−82.9	5000	−33
18	−81.9	6000	−31.25
20	−80.9	7000	−30
25	−79	8000	−28.9
30	−77.25	9000	−27.9
35	−76	10,000	−26.9

mining factors of receiver quality. The use of SAW filters provides sharper selectivity and more linear phase response. These attributes yield better stereo separation and lower distortion. A pulse-counting discriminator also provides low distortion and accurate demodulation of the received signal. Phase-linear low-pass filtering is critical for best stereo separation.

A low-noise RF preamplifier may be added to the system when the received signal level is low. For best performance, the preamplifier should be mounted directly at the receive antenna. Care must be taken, however, to prevent overloading the receiver front-end by unwanted, and often strong, interfering signals.

In areas of frequency congestion, narrowband receivers are important in preventing interference from other transmitters. STL manufacturers have responded to the needs of broadcasters in congested RF environments by providing suitable narrowband STL systems. Such receivers typically incorporate bandpass cavity filters, helical resonators or mechanical coaxial filters. SAW filters and ceramic filters in IF stages also may be included to improve selectivity.

Choose a monaural, composite or digital STL as needed. Compare performance specifications and carefully analyze the long-term benefits of each system. Examine what the STL will need to handle two or five years from now.

Transmission Lines

Figure 4.5-14 shows the primary hardware elements required for an aural STL. Transmission line sections, connections and strain-relief provisions are important for long-term reliability. The main criteria in the selection of transmission line include the following:

- Amount of signal attenuation
- Physical parameters (dielectric material and size)
- Purchase and installation cost.

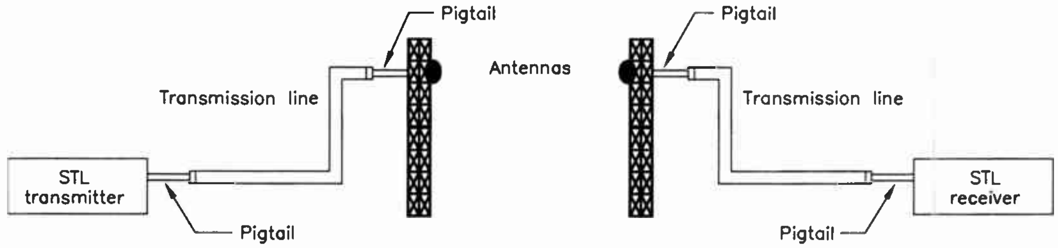


Figure 4.5-14. Principal physical components of an STL system.[2]

In general, the larger the diameter of the transmission line, the lower the loss, and the greater the cost of the line. Loss is also affected by the type of dielectric material used. The most common types of dielectric are air and foam. Air dielectric cable typically requires pressurization and is, therefore, seldom used for 950 MHz installations. For the purpose of gain/loss calculations, cable loss for a particular installation can be determined from the transmission line manufacturer's specification table, given in decibels per unit of measure. Attenuation for common types of coaxial line are given in Table 4.5-5.

Other electrical and mechanical specifications (primarily cable size and impedance) must be compatible with the transmitter, receiver and antennas to be used. Connector loss must also be considered. It is important to minimize each potential source of signal loss when assembling the system. There are no "minor details" in the installation of an STL.

Stain relief must be provided on each end of the cable run (at the transmitter and at the receiver). So-called pigtail or jumper cables are commonly used for this purpose. They permit movement without straining cable and chassis connections. Because the pigtails commonly are terminated with *N*-type male connectors on both ends, the main transmission line must be configured with female *N*-type connectors on both ends if a pair of pigtails are used.

Antenna System

At the frequencies commonly used for STL operation, antennas can readily focus the signal into a narrow beam. This focusing, required for point-to-point communications, provides high gain at both the transmitter and the receiver. Several types of antennas are available, including parabolic and parabolic section. Parabolic antennas are available in solid or grid styles, while parabolic section antennas usually are grid type in design. Antenna models differ in a number of respects, including:

- Gain (directly proportional to size)
- Operating frequency range
- Polarization (most antennas can be set for either horizontal or vertical polarization using universal mounting hardware kits)
- Beamwidth
- Front-to-back ratio
- Wind loading
- Structural strength.

Table 4.5-5
Typical Coaxial Transmission Line Loss at 950 MHz.

Cable Type	Loss (dB/100 Feet)
RG-8/U	8.5
RG-218/U	3.8
1/2-in foam dielectric	3.0
7/8-in foam dielectric	2.0
1/2-in low-loss foam dielectric	2.4
7/8-in low-loss foam dielectric	1.4
1-5/8-in foam dielectric	1.4
1/2-in air dielectric	2.7
7/8-in air dielectric	1.4
1-5/8-in air dielectric	0.7

Environmental factors may also need to be considered so that suitable antenna accessories, such as de-icers and/or radomes, are included.

Antenna gain is specified in decibels referenced to an isotropic antenna (dBi) or decibels referenced to a dipole antenna (dBd). Conversion from dBd to dBi is as follows:

$$\text{dBi} = 2.2 + \text{dBd}$$

For path analysis calculations of system gains and losses, dBi is used.

Figure 4.5-15 plots the response of a common type of STL antenna. Note the high directivity provided by the device and the difference in gain between operation in the 450 MHz band as opposed to the 950 MHz band. Table 4.5-6 lists the typical gain provided by antennas of different physical size.

Mounting Structures

Each antenna requires a mounting structure, such as the side of a roof or building, or an existing or custom-built tower. The structure must be sufficiently high to mount the antenna at the minimum elevation determined by STL path calculations. The structure must also be capable of supporting the weight of the antenna and transmission line, as well as withstanding any forces that affect the antenna tower, such as wind and the weight of snow and ice. Be certain to take into account the mounting hardware used with the antenna, which typically accommodates pipe up to a diameter of 3.5 in. (9 cm). Provisions must be made to secure the transmission line. The structure may also be subject to local building and zoning codes. Strict adherence to these codes is mandatory. Adequate physical clearance must be considered to allow for proper horizontal and vertical alignment during installation.

Hardware Considerations

Depending on the complexity of the STL system, additional hardware may be required. Where more than one transmitter or receiver is used with a single antenna, a combiner, splitter or duplexer will be needed. These items contribute additional loss and must be accounted for in path calculations.

Certain installations may require the installation of an external power amplifier in cases where the output of a standard STL transmitter is insufficient for a particular path. In path calculations, the power output of the external power amplifier (converted to dBm) is substituted for the output of the STL transmitter. In general practice, most engineers choose to use an external power amplifier only as a last resort. Higher gain antennas usually are a more attractive alternative.

SYSTEM PLANNING

The ultimate goal in selecting STL equipment is to choose a combination of STL transmitter, transmission

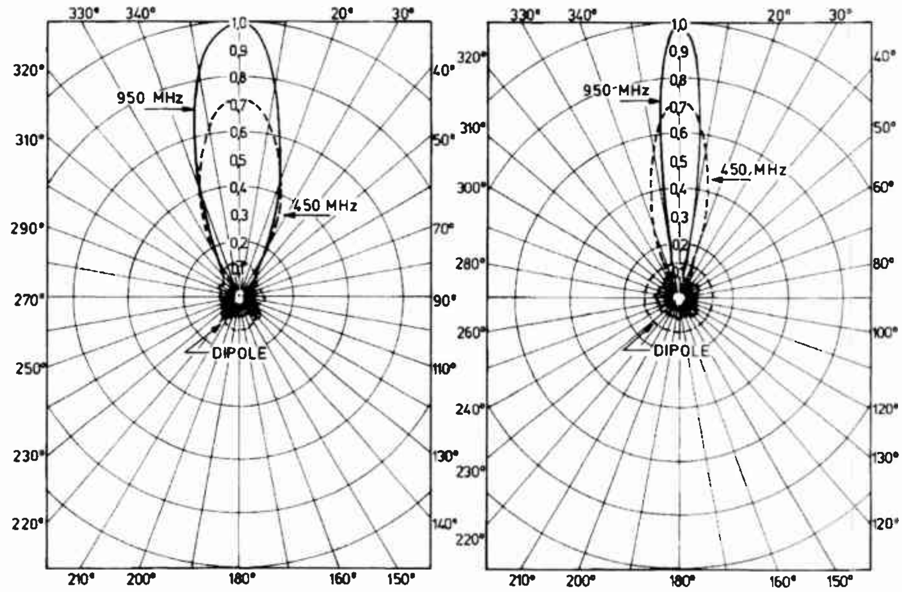


Figure 4.5-15. Radiating plots for a parabolic section antenna at 450 MHz and 950 MHz: (*a*) horizontal response, (*b*) vertical response.[2]

line, antennas and STL receiver that will give adequate quieting (S/N) when transmitting over the path between the studio and the station transmitter. Allowance in the form of *fade margin* must also be made for the uncertainties that the path imposes on the received signal.

The STL transmitter supplies the operating power that drives the system. The transmit and receive antennas add gain to the link. All remaining components add loss. By far the greatest loss results from the path itself. Passive components of the system, including the transmission line, connectors and pigtail cables, also add loss. All gains and losses must be taken into account to ensure that the RF level at the receiver input is adequate. Losses should be minimized wherever possible.

Frequency Selection

Frequency selection would be a relatively simple procedure were it not for other broadcasters, who also may be using STL frequencies and perhaps sharing the same transmitter site. In practice, the choice of operating frequency is governed by the availability of unused STL channels in the area. It is further dictated by the need to avoid interference with other STL users. The local frequency coordinator (under the auspices of the Society of Broadcast Engineers in most areas of the United States) usually can provide information on available frequencies and potential frequency conflicts.

Table 4.5-6
Typical Isotropic Gain for Full Parabolic Antennas at 950 MHz.

Antenna Diameter	Isotropic Gain
4 feet	18.9 dBi
5 feet	21.0 dBi
6 feet	22.0 dBi
8 feet	25.0 dBi
10 feet	27.0 dBi

STL engineering can be divided into three broad categories:

- Frequency selection
- Path layout
- Path gain/loss calculations

Path layout concerns the factors affecting the physical location of the STL transmitter and receiver as well as the intended path. The goal in path layout is to ensure a correctly oriented and unobstructed route for unhampered propagation of the radio wave. Path gain/loss calculations involve analyzing RF power levels from the transmitter output to the receiver input so that an adequate receive level is provided.

Spectrum Considerations

In view of the serious spectrum congestion problems that exist today in many areas of the United States, an STL system should be designed to be as spectrum efficient as possible and—equally important—to be as immune to undesired transmissions as possible. Even if the system will be operated in an area that currently does not have a spectrum congestion problem, there is no guarantee that such a situation will not surface in the near future. In any event, a well-engineered system is also a spectrum-efficient system.

The first rule of spectrum efficiency is to use only the *effective radiated power* (ERP) necessary to do the job. There is no justification for putting 15 W into the air when 5 W will provide the required receiver quieting and fade margin.

A simple and sometimes effective spectrum coordination tool is cross-polarization. Two stations on nearby frequencies may achieve as much as 25 dB RF isolation through the use of different polarizations of transmit antennas, matched by like polarization at their respective receive antennas. Cross-polarization results in varying degrees of success, depending upon the frequency of operation and the surrounding terrain.

Path Layout

Careful path engineering should be performed prior to any licensing work to determine if the proposed locations of the STL transmitter and receiver will be able to achieve the desired results. There is much more to path engineering than simply pointing the transmitting and receiving antennas at each other and turning the equipment on. Antennas cannot always be located in the best possible geographic points because of space availability problems, excessive construction or site rental costs or local/federal licensing difficulties. In such cases, the required path is not the ideal path, and the link will have to be engineered around these fixed points.

Planning for any STL system should begin with an accurate, detailed U.S. Geological Survey (USGS) map covering the proposed path. Note should be made of any natural obstructions (such as mountains, hills or vegetation) or man-made obstructions (such as buildings, water tanks or transmitting towers) in the path.

Line of Sight

Because microwave frequencies are used for STL systems, the signal path is theoretically limited to the line-of-sight between the studio and transmitter locations. In reality, the *radio horizon* is frequently situated beyond the visual horizon. This is the result of the gradual decrease in the refractive index of the atmosphere with increasing altitude above the earth. This effect bends radio waves downward. The degree of bending is characterized by the *K factor*, which is the ratio of the effective earth radius to the true earth radius. A typical value for *K* is 4/3, or 1.33, valid over 90% of the time in most parts of the world. For long paths, consult a map showing the minimum *K* factor occurring in the specific area so that proper antenna heights can be planned. Figure 4.5-16 plots an example path on 4/3 earth graph paper.

STL Site Selection

Usually, there is little choice for the STL transmitter and receiver locations, because most STL systems will be required to provide a path between already determined studio and transmitter points. If, however, a new studio location is being planned, the possible STL path is an important consideration. Figure 4.5-17(a)

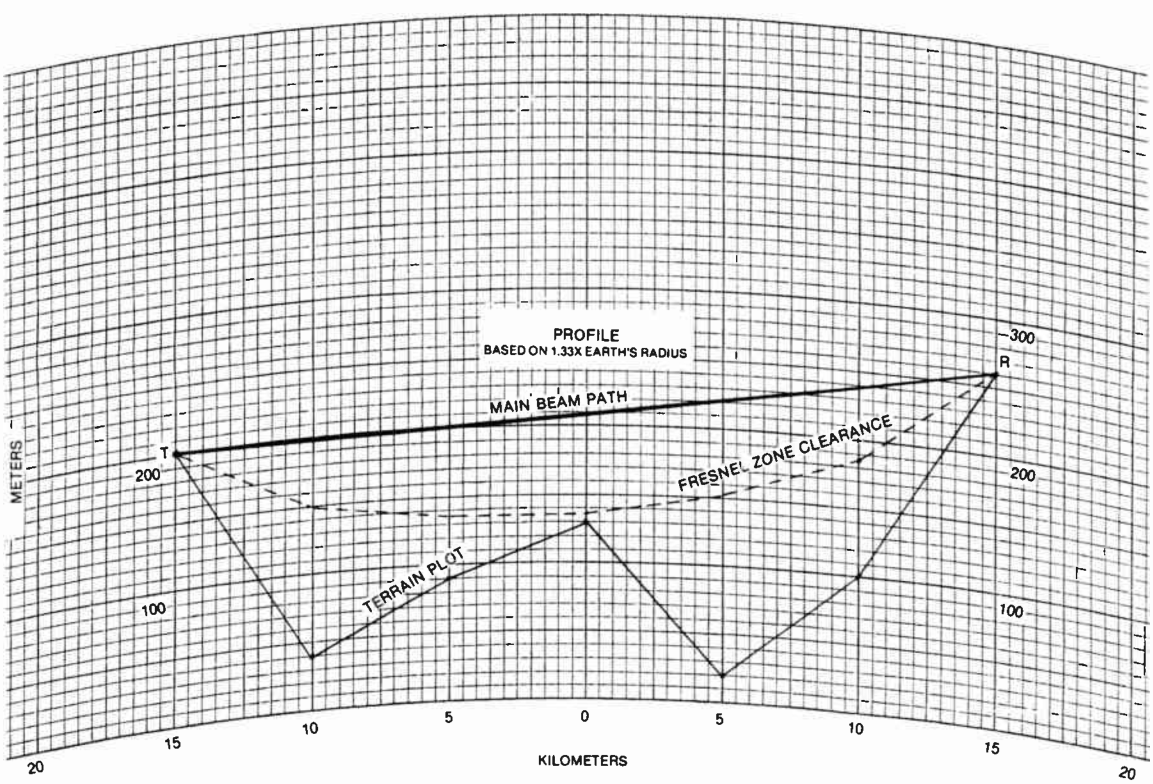


Figure 4.5-16. An aural STL path profile drawn on true-earth radius graph paper.[2]

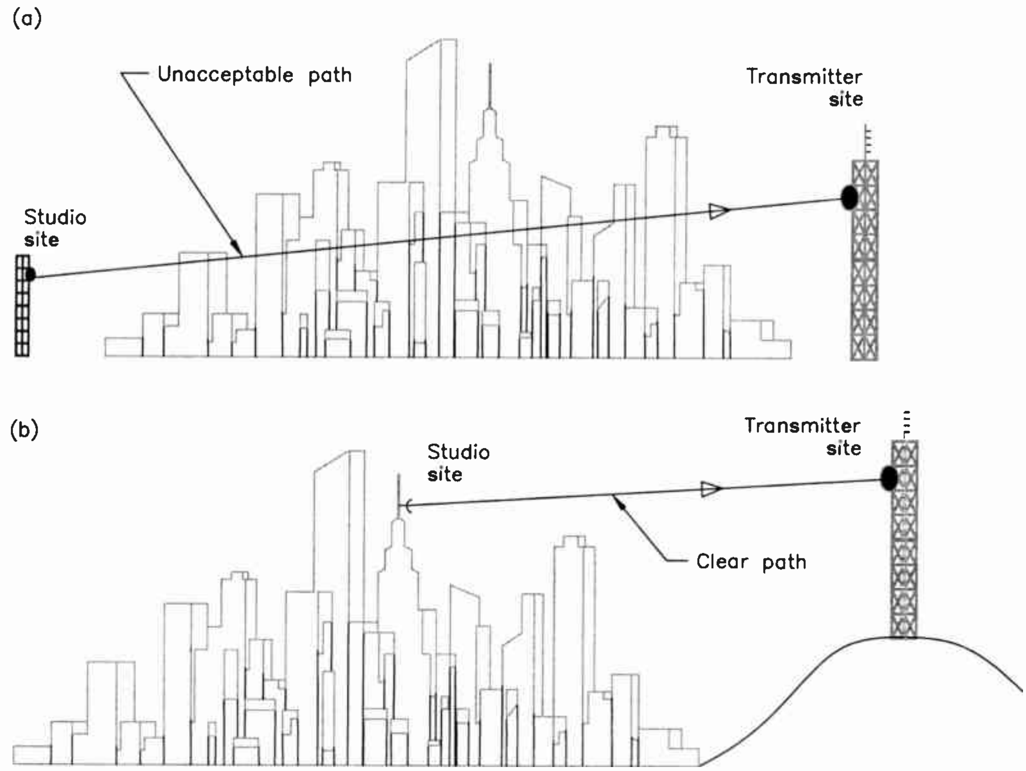


Figure 4.5-17. Path considerations in planning an STL: (*a*) unacceptable path because of obstructions, (*b*) usable path.[2]

illustrates a poor path. Generally speaking, a good path is literally line-of-sight, with no obstructions to block the signal and no other terrestrial or atmospheric conditions that would compromise the path (see Figure 4.5-17(b)).

Terrain Considerations

One of the major tasks required to engineer an STL system is the path analysis between the STL transmitter at the studio and the STL receiver location. To determine what constitutes a clear path, the concept of Fresnel zones for optical theory is applied to radio waves. Most of the electromagnetic energy at a receiving point is concentrated in an elliptical volume that is a function of the distance between the transmit and receive points and the wavelength. The energy outside this volume either cancels or reinforces the energy within the volume, depending on whether the distance that the energy travels to the receive point is longer by an even or odd number of 1/4 wavelengths. Even distances result in radio wave cancellations; odd distances result in radio wave reinforcement (see Figure 4.5-18). The radius of the first Fresnel zone, which defines the boundary of the elliptical volume, is given by the following:

$$F_1 = 72.1\sqrt{\frac{d_1 d_2}{fD}}$$

Where:

F_1 = first Fresnel zone radius in feet
d_1 = distance from the transmitting antenna to the obstruction in miles
D = total path length in miles
d_2 = $D - d_1$ in miles
f = frequency in GHz
H = distance from the top of the obstruction to the radio path

For reliable operation, obstructions should not project into the area thus defined. Empirical studies, however, have shown that performance is substantially the same for H greater than 0.6 F_1.

The first step in evaluating the path is to make a subjective check of the planned route. First, determine that a reasonable possibility of success exists. Next, draw a best-case route for the path. If obstructions are found, consider accomplishing the path by using a repeater system or by shifting the location of the STL transmit antenna. Although a detailed path analysis may not be required in cases where line-of-sight is clearly present and the distance is less than about 10 miles (16.1 km), it is still good engineering practice to review some of the basic elements of the evaluation.

Obtain an accurate map showing the topography between the STL transmitter and receiver locations. After determining the transmitter and receiver sites on

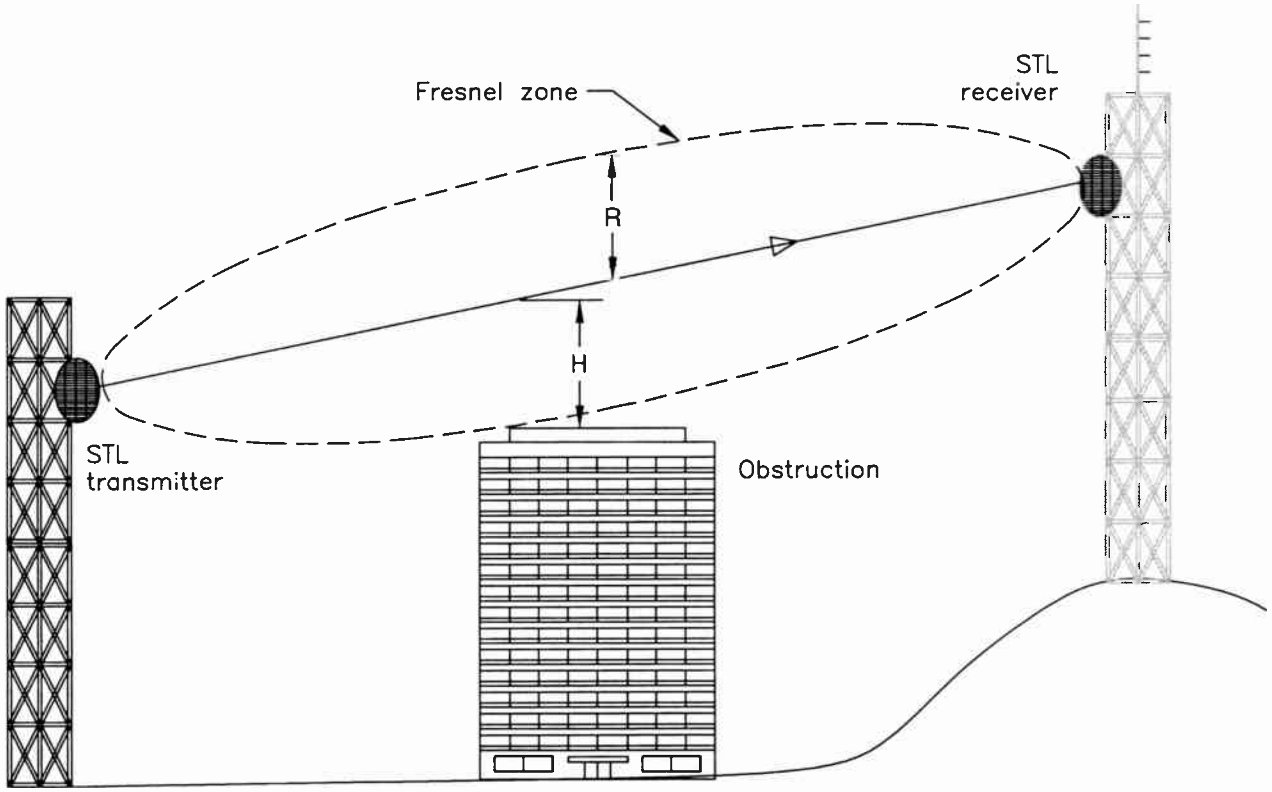


Figure 4.5-18. Fresnel zone clearance for an STL path.[2]

the map, connect them with a straight line showing the proposed path. After the path has been drawn, a protractor can be used to determine the direction (azimuth) of the path in degrees from true North. This data will later assist in antenna setup and is necessary for filling out the appropriate FCC application.

Using the scale of miles on the map, measure the distance between the transmitter and receiver sites. Determine the altitude of the proposed transmit antenna location from the contour lines on the map and add to that the height of the building or other structure on which the antenna will be mounted. Make a similar determination for the receive antenna location. Adjust the heights to ensure that all obstructions are cleared. Depending on the path length and the height of the antennas, it may be necessary to take the curvature of the earth into account. If so, use true-earth radius graph paper to plot a cross-section of the path.

Study the map to see what terrain features are between the path points. Prior to making any other evaluations, conduct a visual field survey. Check for any structures or features not listed on the map. Anticipate any possible new construction or tree growth that may cause problems in the future.

The terrain from the transmitting antenna to the receiving antenna must be examined not only for obstructions, but for reflection possibilities as well. A large body of water will usually cause problems for an STL system. If the water is an even number of Fresnel zones from the direct path, signal attenuation will likely occur at the receiver. Temperature changes and tidal conditions will also have an effect. Likewise, thick vegetation or forested areas can be reflective to RF signals when wet, creating a similar (but not so troublesome) problem. Generally, the solution to reflection conditions is to change either the transmitting or receiving antenna height. In extreme cases, a diversity reception system may also be used.

Path Reliability

The long term reliability of an STL path is determined in large part by the initial engineering work done before the system is installed. Figure 4.5-19 charts gains and losses through a typical system. The most important factors are free space loss and allowance for fade margin. Thereafter, effects such as diffraction, reflection, refraction, absorption, scattering and terrain loss must be considered.

A gain and loss balance sheet should be computed to determine the fade margin of the planned STL system. An adequate fade margin is vital to reliable performance because a link that is operating on the edge of the minimum acceptable receiver quieting will likely encounter problems at some point. Normal component aging in the receiver or transmitter can cause a loss in received signal level and, thus, degrade system performance. Atmospheric conditions, such as severe weather in the area or ice on the transmitting or receiv-

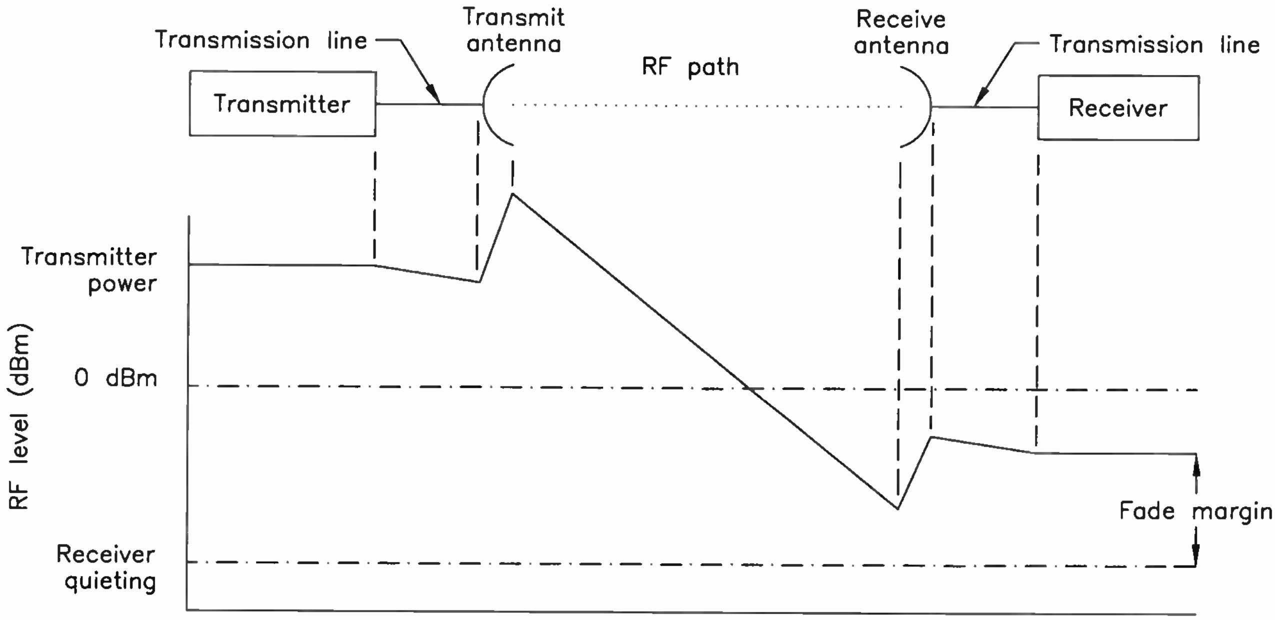


Figure 4.5-19. Path analysis of plotting gains vs. losses for an STL.[2]

ing antennas, can also cause sharp fading and even a complete loss of signal if an adequate fade margin above minimum receiver quieting is not provided. The STL fade margin can be computed using the following equations:

$$G_s = G_t + G_{ta} + G_{ra}$$

Where:

G_s = total system gain in decibels
G_t = transmitter power output in dBm
G_{ta} = transmit antenna gain in dBi
G_{ra} = receive antenna gain in dBi

The values for G_{ta} and G_{ra} are gathered from the antenna manufacturer's literature.

The value for G_t is given by the following:

$$G_t = 30 + 10 \log P_o$$

Where:

G_t = transmitter power output in dBm
P_o = transmitter power output in watts

The total system losses are then computed:

$$L_s = L_p + L_l + L_c + L_m$$

Where:

L_s = total system losses in decibels
L_p = path loss in dB
L_l = transmission line loss in dB
L_c = connector losses in dB
L_m = miscellaneous losses in dB

The values for L_l and L_c can be determined from manufacturer's literature. Figure 4.5-20 shows typical loss values for 0.5 in. (1 cm) foam-filled transmission line. A reasonable value for connector loss with components normally used in 0.5 in. (1 cm) coax installations is 0.5 dB. The value for L_p can be determined by using the following formula:

$$L_p = 36.6 + 20 \log F + 20 \log D$$

Where:

L_p = free space attenuation loss between two isotropic radiators (in dB)
F = frequency of operation in megahertz
D = distance between the antennas in statute miles

Free space loss can also be found using a table of approximate values, as given in Table 4.5-7.

With the foregoing information available, the fade margin can then be calculated:

$$M_f = G_s - L_s - R_m$$

Where:

M_f = fade margin (in dB)
G_s = total system gain (dB)
L_s = total system losses (dB)
R_m = minimum signal strength required for the target S/N in dBm (a negative number)

G_s and L_s are determined by the equations given previously. R_m (receiver sensitivity) is determined from the receiver manufacturer's specifications. If the manufacturer gives a receiver sensitivity figure in microvolts, the following formula can be used to convert to dBm:

$$R_m = 20 \log\left(\frac{V_r \times 10^{-6}}{0.7746}\right)$$

Where:

R_m = minimum required signal strength (in dBm)
V_r = receiver sensitivity (in μV)

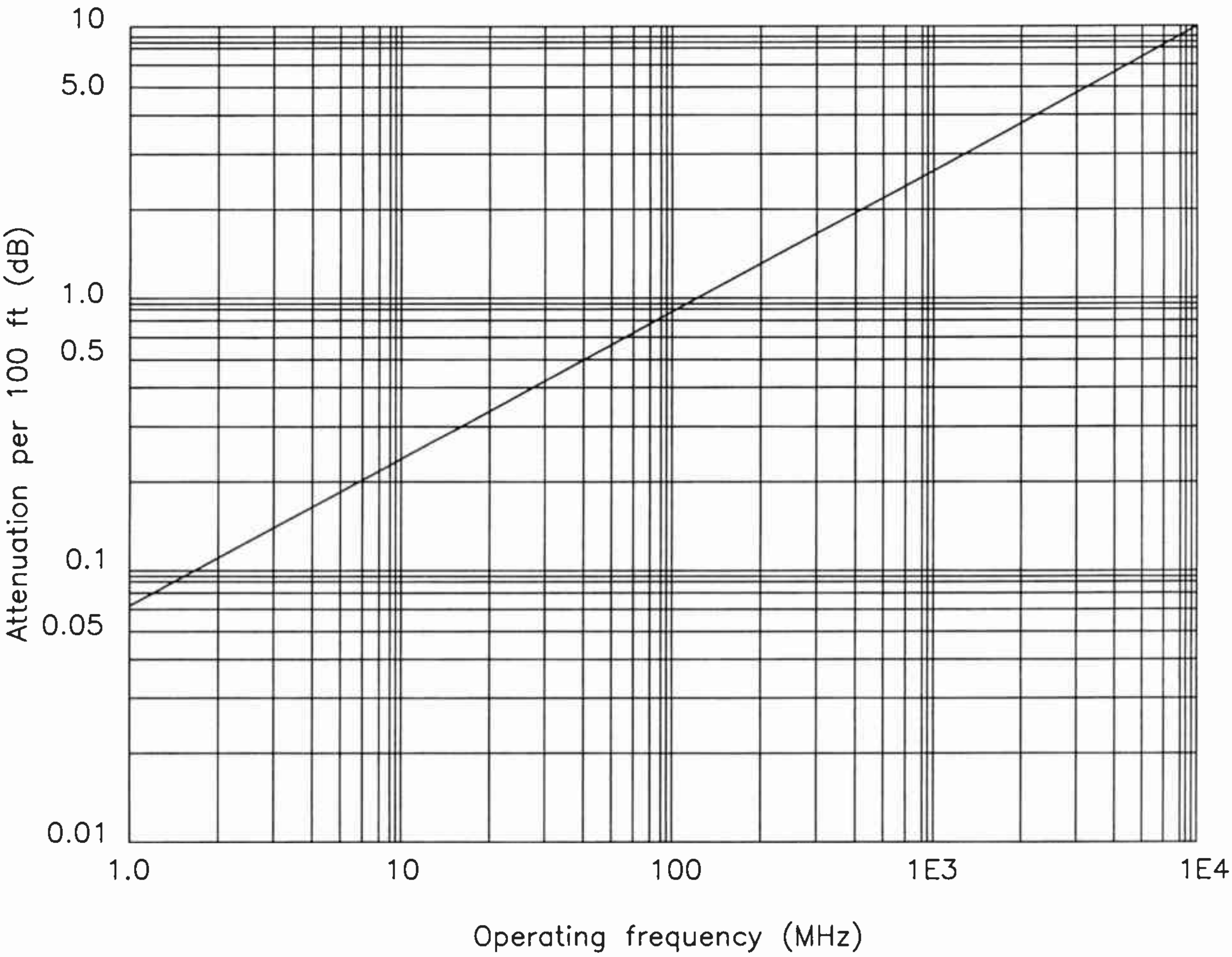


Figure 4.5-20. Loss vs. frequency for 1/2 in. foam dielectric transmission line.[2]

In order to predict accurately the performance of the STL radio link, the value of R_m must be determined carefully. For maximum system performance and reliability, the fade margin determination should be made based upon the signal level required to provide the minimum acceptable receiver S/N performance. Longer paths require greater margins for fade.

Table 4.5-7
Path Attenuation at 950 MHz.

Distance in Miles	Loss in dB	Distance in Miles	Loss in dB
1	−96.2	16	−120.2
2	−102.2	17	−120.7
3	−105.7	18	−121.3
4	−108.2	19	−121.7
5	−110.1	20	−122.2
6	−111.7	21	−122.6
7	−113.1	22	−123.0
8	−114.2	23	−123.4
9	−115.3	24	−123.8
10	−116.2	25	−124.1
11	−116.9	26	−124.5
12	−117.7	27	−124.8
13	−118.4	28	−125.1
14	−119.1	29	−125.4
15	−119.7	30	−125.7

The primary cause of signal fade in an STL system below 1.0 GHz is changes in the refractive indexes of the atmosphere along the signal path. These fluctuations affect the path of reflected or refracted signals differently from the direct, line-of-sight signal. When the interfering signals combine with the direct signal, the level at the receiver increases or decreases depending upon the degree of reinforcement or cancellation. Because atmospheric conditions are seldom stable, some fade margin is always necessary.

Another cause of signal fade is *earth bulge* (or *inverse beam*) fading, where the overall refractive index in the vicinity of the signal path decreases, thus hindering the full signal from reaching the receive antenna. Again, allowance for signal fade will minimize degradation. Precipitation is another potential cause of signal fading, although it is not generally considered significant at frequencies below 1.0 GHz.

Approximate fade margin also can be determined from Table 4.5-8. The relationship between system reliability and fade margin is detailed in Table 4.5-9.

Table 4.5-8

Recommended Fade Margin as a Function of Path Length.

Path Length	Fade Margin
5 mile	5 dB
10 mile	7 dB
15 mile	15 dB
20 mile	22 dB
25 mile	27 dB
30 mile	30 dB

Dealing with Problem Paths

The STL systems described thus far have been simple installations with a transmitter at the studio and a receiver at the broadcast transmitter site. There are cases, however, where a simple, single-hop STL is not be practical. Reasons include the following:

- The distance is too great to be reasonably covered by a single-hop system
- A direct path cannot be used because of obstructions of some type
- An unusually large fade margin is required for the application.

Figure 4.5-21 shows three instances where a two-hop system is required. A two-hop repeater uses two transmitters and two receivers, with one pair located at the intermediate site. As shown in Figure 4.5-21(c), interference avoidance is another practical application of a multi-hop system.

With a conventional active repeater, the received signal is demodulated down to the baseband at the intermediate site before being retransmitted. This approach is shown in Figure 4.5-22. Unfortunately, this method results in inevitable degradation of the analog signal. A better approach, an IF repeater, is shown in Figure 4.5-23. Instead of being demodulated to the baseband level, the received signal is brought down to the IF level before being injected into the repeater transmitter.

Hot-Standby Redundant Systems

The importance of continuous, uninterrupted service from an STL system cannot be overstated. Modern STL equipment is stable and reliable; failures rarely occur. However, a station may elect to install a hot-standby redundant system to ensure zero downtime from the STL. Figure 4.5-24 shows a block diagram of a common hot-standby system. The installation includes a second transmitter and a second receiver. An automatic changeover unit is included for switching purposes. The changeover system monitors each transmitter and receiver pair to sense critical parameters. If a failure occurs, the changeover controller switches from the faulty transmitter to the standby transmitter, or from the faulty receiver to the standby receiver. The changeover units work independently.

This system can be simplified to include only hot-standby provisions at the transmitter site or at the receiver site. If this approach is taken, a case can be

Table 4.5-9

Relationship between Fade Margin, Reliability and Outage Time for Rayleigh Distributed Paths.

Fade Margin (dB)	Path Reliability/ Availability	Outage Hours per Year	Outage Minutes per Month	Outage Seconds per Day
10	90.4837	834.20	4170.98	8222.05
20	99.0050	87.22	436.12	859.69
21	99.2088	69.35	346.77	683.58
22	99.3710	55.14	275.68	543.43
23	99.5001	43.82	219.12	431.94
24	99.6027	34.83	174.14	343.28
25	99.6843	27.68	138.38	272.79
26	99.7491	21.99	109.96	216.75
27	99.9007	17.47	87.37	172.22
28	99.8416	13.88	69.41	136.83
29	99.8742	11.03	55.14	108.70
30	99.9000	8.76	43.81	86.36
31	99.9206	6.96	34.80	68.60
32	99.9369	5.53	27.65	54.50
33	99.9499	4.39	21.96	43.29
34	99.9602	3.49	17.45	34.39
35	99.9684	2.77	13.86	27.32
36	99.9749	2.20	11.01	21.70
37	99.9800	1.75	8.74	17.24
38	99.9842	1.39	6.95	13.69
39	99.9874	1.10	5.52	10.88
40	99.9900	0.88	4.38	8.64
41	99.9921	0.70	3.48	6.86
42	99.9937	0.55	2.77	5.45
43	99.9950	0.44	2.20	4.33
44	99.9960	0.35	1.74	3.44
45	99.9968	0.28	1.39	2.73
50	99.9990	0.09	0.44	0.86
55	99.9997	0.03	0.14	0.27
60	99.9999	0.01	0.04	0.09

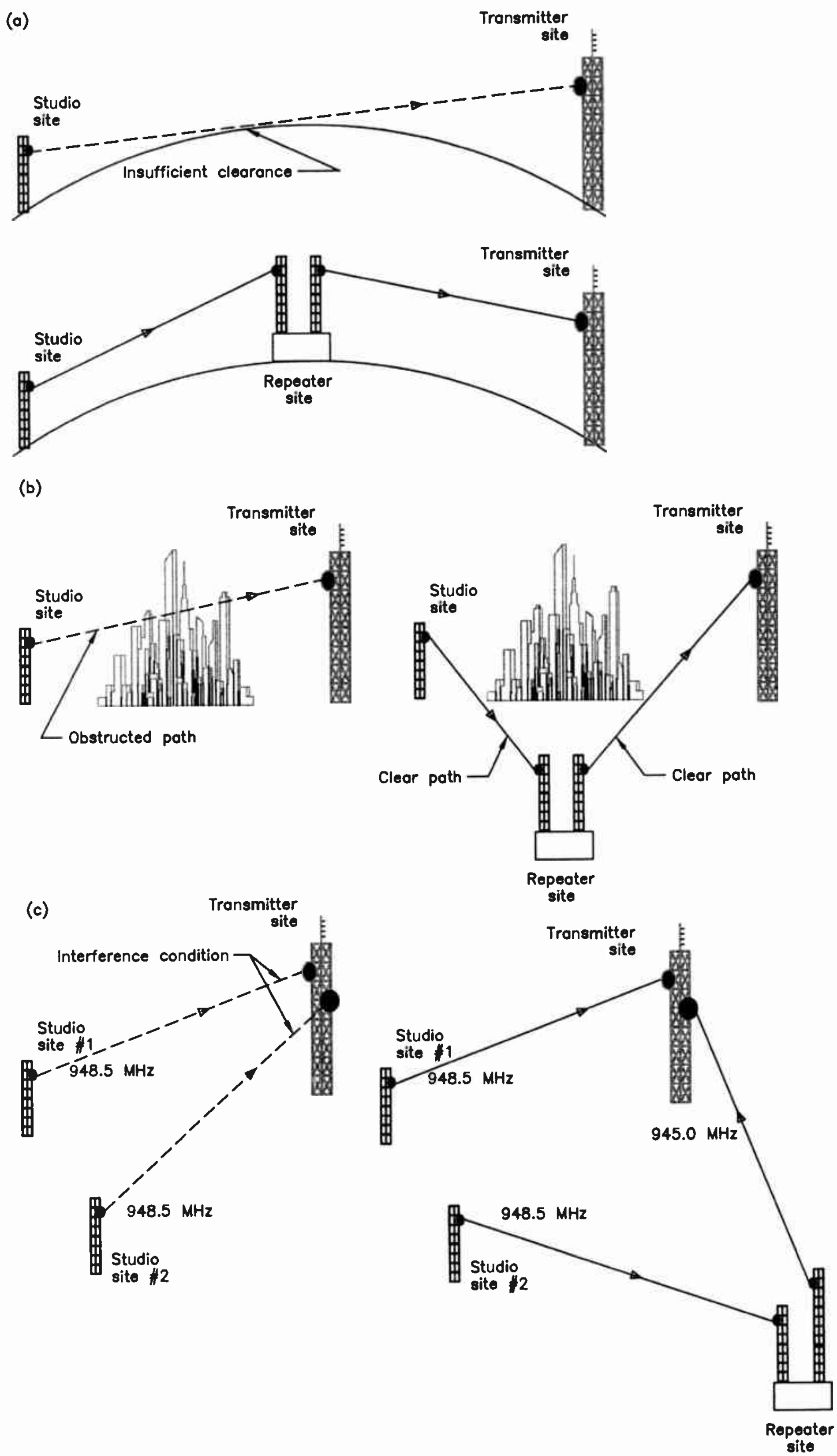


Figure 4.5-21. Applications requiring the use of a multi-hop STL system: (*a*) excessively long path, (*b*) path obstructions, (*c*) interference concerns.[2]

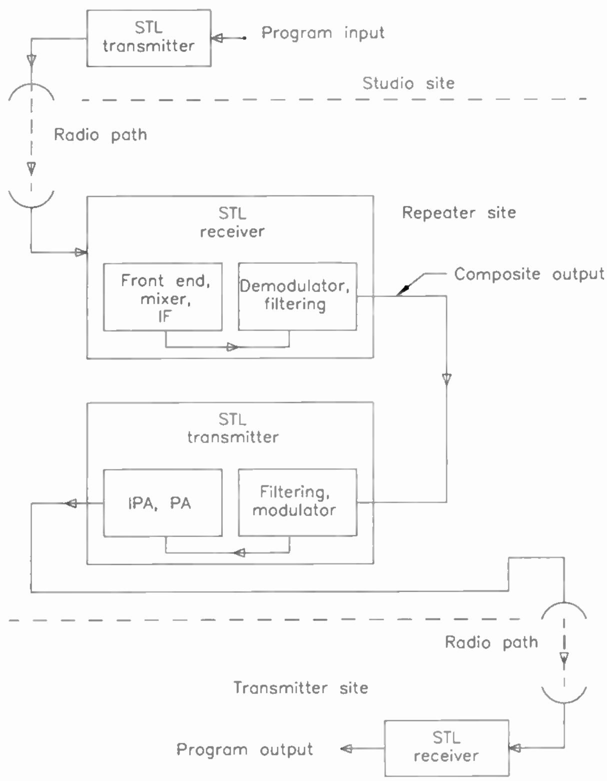


Figure 4.5-22. Repeater link using conventional demodulation and remodulation of the composite signal.[2]

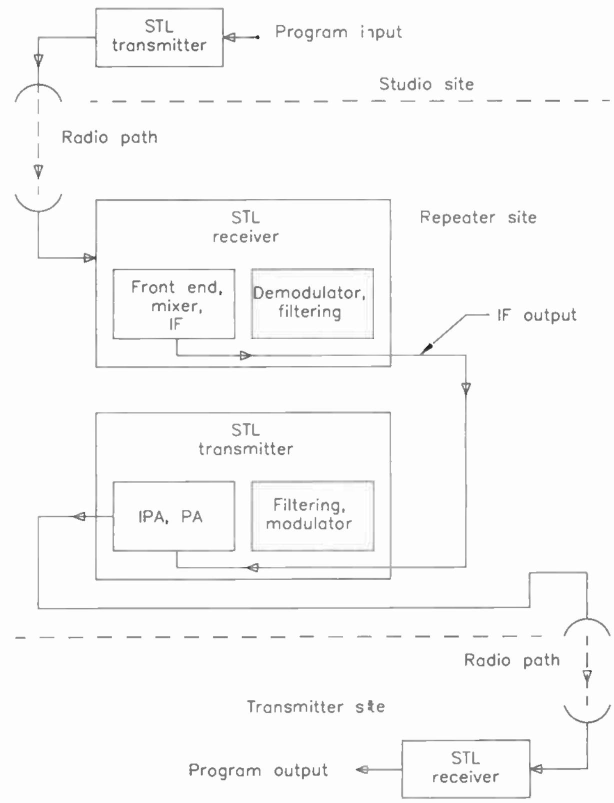


Figure 4.5-23. Repeater link using IF transfer of the composite signal.[2]

made for either backup system. One argument states that the transmitter is more likely to fail because it contains high-power stages, which typically are more prone to problems than low-power circuits. On the other hand, the transmitter is almost always more accessible for repair—being at the studio site—than the receiver.

Pre-installation Checkout

The best planned STL system, well engineered and comprised of the finest equipment, will not perform properly if incorrectly installed. Quality control steps taken before installation actually begins is an important part of the installation process. Delivery schedules for STL hardware vary widely. After all the equipment is ordered, components will arrive, typically, over a period of some weeks. Good engineering practice dictates that each component is checked or inspected as it arrives. This process is preferable to waiting for all of the components to arrive on site because any problems discovered at that point will delay the project.

The STL transmitter can be operated into a dummy load to confirm proper operation. With an ideal load, the transmitter front panel readings should correspond closely with the final test sheet supplied by the manufacturer. Modulation can be applied to the transmitter to ensure proper operation.

The STL receiver is more difficult to test effectively with the instruments typically available at a radio station. The engineer can, however, bench test the receiver by operating the transmitter into a dummy load and attaching a short wire to the receiver antenna input. Be sure to use the proper connector on the receiver to avoid possible damage to the center pin. Maintain sufficient separation of the transmitter and receiver to prevent overloading the receiver front-end. Arranged in this manner, the STL system can be checked from end-to-end with a signal generator (or program material) on the bench.

For a digital STL, the transmitter, receiver and modem/multiplexer can be tested in a similar way. Bench tests before installation usually resolve hidden or unexpected problems associated with equipment operation and/or interfacing. Functionality of the transmitter and receiver should be checked before adding the encoder to the system. In this manner, any problems detected can be readily pinpointed.

An STL using one or more subcarriers for program relay, closed-circuit communications or remote control functions can be bench tested in a similar manner. Having all hardware on one bench and easily accessible makes adjustment of levels and other parameters much easier than when the units are separated by many miles.

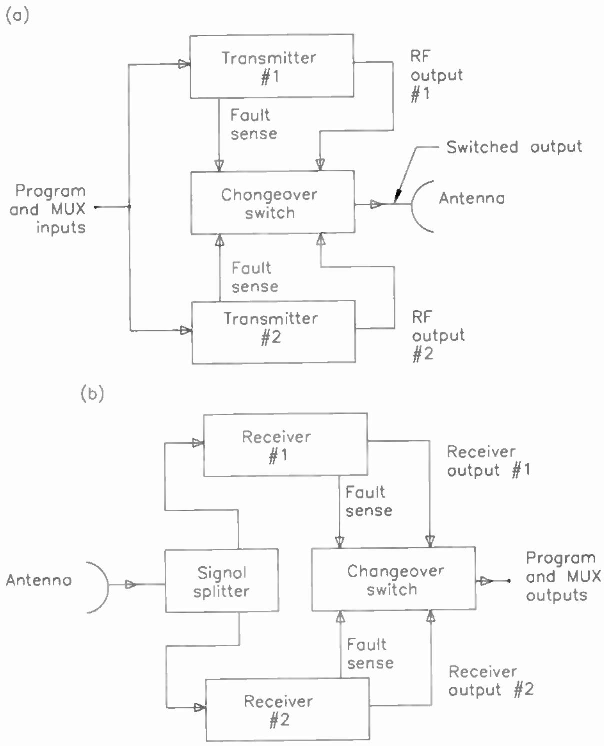


Figure 4.5-24. Equipment configuration for a hot-standby system: (*a*) studio site, (*b*) broadcast transmitter site.[2]

Nearly all STL systems checked in this manner pass specification testing with no problem. Still, there may be instances where a unit is damaged during shipment. It is far easier to solve such problems before the hardware is installed.

Antennas should be given a close visual inspection. Many antenna models used for STL work are shipped partially assembled. Final assembly of the antennas should be completed before the installation process is begun. Pay particular attention to mounting brackets; make sure the hardware will mate properly with the tower or structure on which the antennas will be mounted. Check carefully for missing parts.

The transmission line and connectors require no pre-installation quality control however, it is suggested that as many connectors as possible be installed in the shop, rather than at the site. *N*-type connectors require patience and skill to install. If possible, have the cable supplier attach the proper connectors to both ends of the cable reel. In this way, only two connectors will need to be installed in the field. Always place the manufacturer-installed connector ends at the antennas. This provides the best assurance of trouble-free operation in an exposed environment. Consider ordering a couple of extra connectors just in case a part is lost or damaged during construction. The engineer may also want to use one connector as a test case to become more familiar with the required installation technique.

The test equipment required for pre-installation checkout is usually readily available in the radio station shop. Basic items include the following:

- A high quality 50 Ω dummy load capable of dissipating approximately 25 W
- An in-line RF power output meter capable of reading forward and reverse power at 1.0 GHz
- Audio frequency signal generator
- Audio frequency distortion analyzer
- Frequency counter accurate to 1.0 GHz.

If possible, a spectrum analyzer also should be available. Although most radio stations do not own a spectrum analyzer, engineers in a given market sometimes rent one to be shared among the stations for a period of several weeks.

Figure 4.5-25 shows a typical bench test setup for an STL system.

Planning Cable Runs

Routing of the transmission line from the transmitter to the antenna is an often overlooked element of STL installation. Confirm that there is a practical route for the coax from the transmitter to the rooftop or tower location. Double-check the length of the transmission line path at both the studio site and the station transmitter site. If the actual run is longer than predicted during fade margin calculations, run the numbers again to be certain that a problem will not develop during system commissioning. Confirm also that the pigtail cables to be used are sufficiently long and have the proper connector terminations.

Installation

The locations commonly used for broadcast transmitter sites are seldom ideal from an environmental standpoint. The site may be difficult to reach during certain parts of the year, very hot during the summer and very cold during the winter. For these reasons, rugged equipment should be chosen and properly installed. Temperature extremes can cause problems for frequency-determining elements, as well as accessories such as cavity filters, preselectors and preamplifiers. Environmental control at the broadcast transmitter site, therefore, is highly recommended.

The STL transmitter and receiver should be mounted in an equipment rack in a protected location adjacent to the stereo generator at the studio site and adjacent to the exciter at the broadcast transmitter site. Keep all cable runs as short and direct as possible. Excessive cable lengths between the stereo generator and the STL transmitter, or between the STL receiver and the exciter, can degrade stereo separation and frequency response. Follow good grounding practices at all times.

The antenna presents probably the greatest installation challenge. Because of its directional nature, the antenna must be properly oriented. Compass bearings are desirable along with whatever visual sightings are practical. The received signal strength at the broadcast transmitter site can be used as an indication of proper orientation of the STL receive antenna. Ensure also

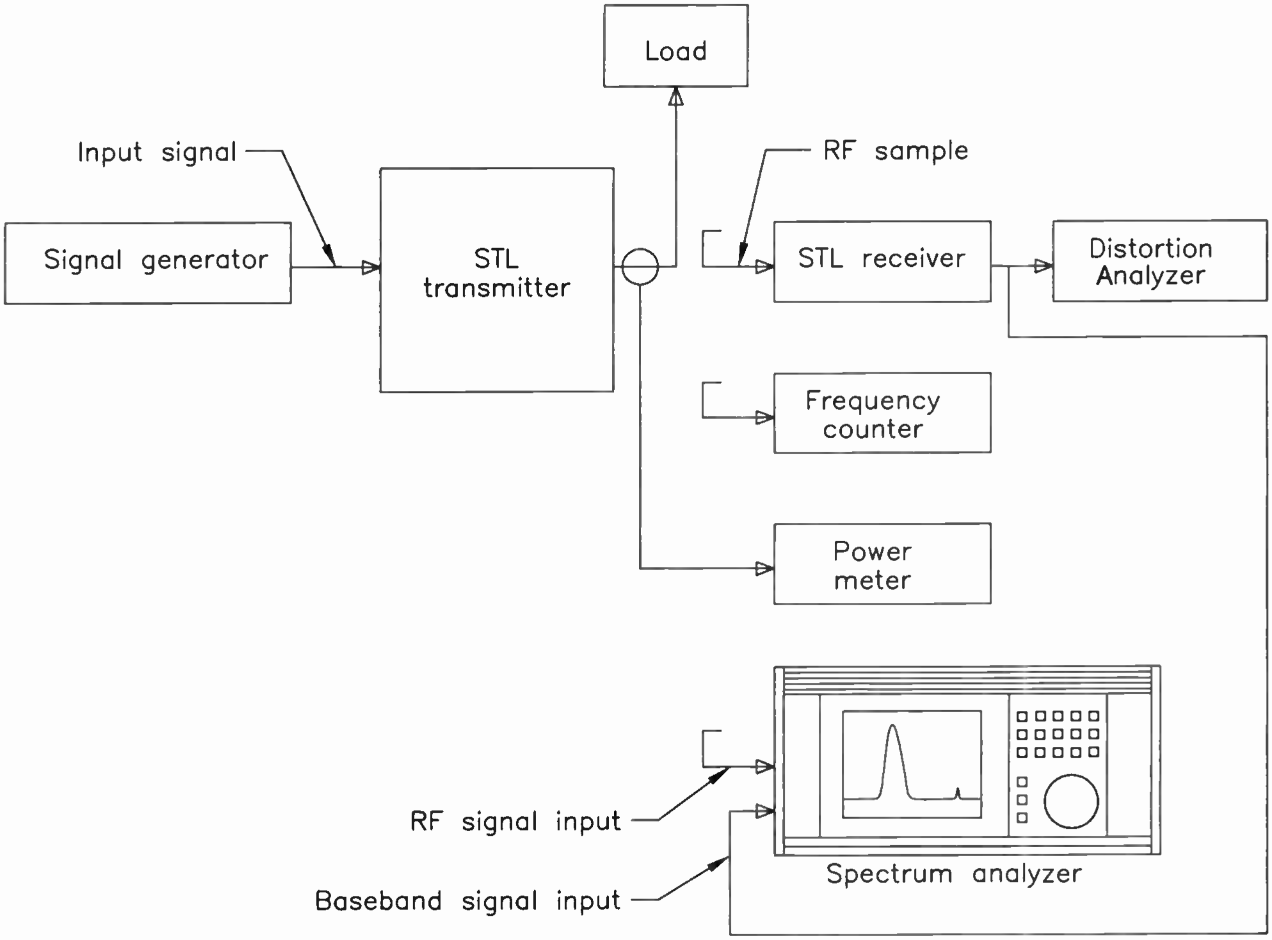


Figure 4.5-25. Equipment setup for bench testing an STL system prior to installation.[2]

that the antenna is set to the proper angle relative to the horizon. Because the STL antennas may be located high on a tower or building, a professional tower crew may be required to mount the devices. Make certain that all cables on the tower are securely fastened to prevent stress and wear. Seal the external connections with sealant made for that purpose and completely wrap the connection joints with tape.

While it is desirable to mount the receive antenna high on a tower to ensure a good path, it is also good engineering practice to keep the antenna as far as possible from radiating elements on the structure, such as the station's main broadcast antenna. Other potential sources of RF problems include two-way radio, cellular radio and TSL systems. If the STL receive antenna is located in a strong RF field, the front-end may be overloaded or *desensitized*. Placing the STL antenna close to other antennas on the structure also can detune the antennas, degrading the performance of both systems.

One of the most common problems encountered during installation is damage to the transmission line. Lines must not be twisted or bent beyond their minimum bending radius. To retain its characteristic impedance, the line must be treated with care. A line damaged by carelessness can cause poor performance of the system.

The transmission line must be properly grounded. As illustrated in Figure 4.5-26, grounding typically is performed at the point where the line leaves the tower and at the point where it enters the building. This procedure, obviously, applies only to grounded tower systems. Grounding kits are available for this purpose.

In situations where the STL antenna is to be mounted on an AM antenna tower, which typically is insulated from ground, an *isocoupler* will need to be installed where the line begins its run up the tower. The isocoupler will pass the STL signal efficiently, while providing a high impedance to ground for the tower at the AM frequency. When this is done, the base impedance of the tower should be expected to change slightly.

Checking System Operation

After all of the components of the STL system have been installed, checkout can begin. Begin by powering

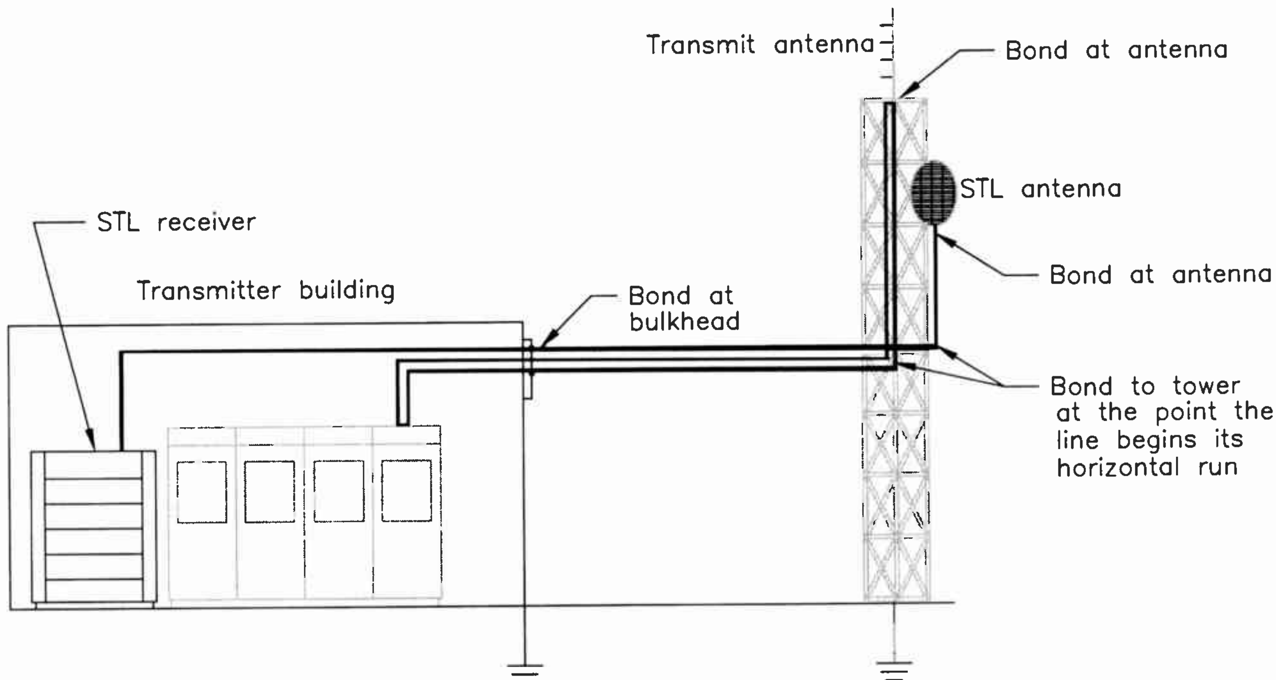


Figure 4.5-26. Grounding practice for an STL transmission line on a tower.[2]

up the STL transmitter. Confirm that all meter readings are within normal parameters. If any problems are detected, shut the system down immediately and troubleshoot the failure.

After proper operation has been confirmed, adjust the modulation levels into the STL transmitter according to the instructions given with the unit. If a digital modem/multiplexer is used, make the necessary adjustments on that unit.

Let the transmitter run for a period of time—several hours or so—and check the operating parameters regularly to be certain that no problems develop. Pay particular attention to voltage standing wave radio (VSWR) and internal high-temperature alarms. If abnormal operation is suspected, consult the manufacturer's equipment manual.

After checkout of the STL transmitter is complete, power up the STL receiver and confirm correct parameters on the front-panel meter. Check and adjust program levels as needed. If a digital modem/multiplexer is used in the installation, check the status monitors on the unit. If displayed, observe the bit error rate monitor to be sure the link is operating properly.

Using an accurate modulation monitor at the broadcast transmitter site, adjust the modulation level of the transmitter. Check operation of the STL system carefully for an hour or so before committing the system to the air.

After the link is operating properly, run a complete proof-of-performance on the broadcast chain. Document the measured parameters so they can be used as a benchmark for later reference.

Troubleshooting

The most common problem associated with STL commissioning is high VSWR. If the indicated VSWR is outside normal limits, the transmission line, a connector or an antenna usually is at fault. A dummy load may be used to determine where the fault lies. Begin at the first pigtail and substitute the dummy load for the transmission line. At each point the load is substituted for the line and antenna, check the displayed VSWR on the transmitter. It is necessary, of course, to power down the transmitter while the dummy load is moved from one point to the next.

Operating the System

Regular performance tests should be made on the STL system. Routine inspections allow the engineer to spot potential problems that could cause a system failure if left unattended.

The STL transmitter and receiver should provide trouble-free performance for many years. If problems develop, the likely trouble areas will be the antennas and transmission lines—the components exposed to the elements. The best check of system performance at the studio is the STL transmitter multimeter. A periodic check of the meter will help to spot oncoming problems. Readings that change gradually can provide clues that will help prevent a failure. Many users will wish to monitor forward power continuously. Reverse power also should be checked regularly to monitor the condition of the line and antennas, as discussed previously. Compare the test data taken after system commissioning with the operating parameters observed during routine checks. They should closely agree.

Operation of the receiver is best checked with its built-in multimeter. The RF level indication should be carefully noted during initial installation and subsequent observations should be compared with this reference. Remember that there will be some change in signal strength readings because of weather conditions and temperature variations. Consider the fade margin conditions used in the path analysis calculations when making judgments about the observed readings. Be aware of unusual variations resulting from temperature inversions, if they are common in the area. If trouble is experienced with the receiver, the possibility of interference from another STL or another service should not be overlooked. A spectrum analyzer is invaluable for such troubleshooting.

For digital STL systems, error rate monitors are provided to assess overall performance. As long as the status indicators on the front panel indicate proper operation, the system usually can be considered to be transparent to the program material.

The most definitive overall check of the system will be the audio proof-of-performance. Primary attention should be given to the noise measurement. If this is not satisfactory, it will be impossible to achieve meaningful distortion measurements because noise will be indicated as distortion by the analyzer. For dual monaural systems, the engineer will need to carefully check left and right channel balance.

A broadcast system is only as strong as its weakest component. Before placing blame for poor performance on the STL, start at the beginning of the broadcast chain and follow the signal step-by-step to the point where it deteriorates.

REFERENCES

1. Rollins, William W., and Robert L. Band, "T1 Digital STL: Discrete vs. Composite Transmission," *NAB 1996 Broadcast Engineering Conference Proceedings,* National Association of Broadcasters, Washington, DC, pp. 356–359, 1996.
2. Whitaker, Jerry C., ed., *A Primer: Digital Aural Studio to Transmitter Links,* TFT, Santa Clara, CA, 1994.

BIBLIOGRAPHY

Hauptstuek, Jim, "Interconnecting the Digital Chain," *NAB 1996 Broadcast Engineering Conference Proceedings,* National Association of Broadcasters, Washington, DC, pp. 360–358, 1996.

McClanahan, M. E., "Aural Broadcast Auxiliary Links," *NAB Engineering Handbook*, 8th ed., E. B. Crutchfield, National Association of Broadcasters, Washington, DC, pp. 671–678, 1992.

Parker, Darryl, "TFT DMM92 Meets STL Requirements," *Radio World,* Falls Church, VA, October 21, 1992.

Salek, Stanley, "Analysis of FM Booster System Configurations," *Proceedings of the 1992 NAB Broadcast Engineering Conference*, National Association of Broadcasters, Washington, DC, April 1992.

Whitaker, Jerry C., and Skip Pizzi, "Radio Electronic News Gathering and Field Production," *NAB Engineering Handbook,* 8th ed., E. B. Crutchfield, National Association of Broadcasters, Washington, DC, pp. 1051–1072, 1992.

4.6 RADIO BROADCAST DATA SYSTEM (RBDS)

SCOTT A. WRIGHT
DELPHI DELCO ELECTRONICS SYSTEMS, KOKOMO, IN

INTRODUCTION

Since the adoption of the Radio Broadcast Data System (RBDS) standard in 1993, much debate has surrounded the success of this system. RBDS is the classical *chicken and egg* problem whereby broadcasters debate purchasing equipment when so few receivers are available, and receiver manufacturers debate introducing capable receivers with so few broadcasters on the air. However, the tides have turned recently and broadcasters have little to fear about making the decision to broadcast RBDS. Currently, RDS is standard or optional in automobiles from General Motors, Ford, Porsche, Audi, Volvo, Jaguar, SAAB, and Mazda. RDS receivers are also available for home, car aftermarket, portable and personal computer use, offered by manufacturers such as Denon, Pioneer, Kenwood, Blaupunkt, Becker, Sangean, Onkyo, and ADS Technologies. While RDS is being introduced into premium systems, its application in many new models is a given *if* more broadcasters get on the air. The risk is low with broadcast encoders available starting at under $400. As of this writing, broadcast equipment manufacturers include RE America, Inovonics, Belar, Rohde and Schwartz, Modulation Sciences and Circuit Research Labs. RDS equipment is available through a number of broadcast supply houses as well, and other manufacturers may have, or be planning to offer, RDS equipment. RDS certainly is accepted by consumers in Europe, where it was introduced in 1982. Since that time over 50 million receivers have been sold, and over 500 different models are currently available.

THE RBDS AND RDS STANDARDS

The Radio Broadcast Data System (RBDS) Standard was adopted in January 1993. It was produced by the National Radio Systems Committee (NRSC) which is a joint committee of the National Association of Broadcasters (NAB) and the Consumer Electronics Manufacturers Association (CEMA). Thus, RBDS is an agreed upon standard between broadcasters and receiver manufacturers. The U.S. RBDS Standard is largely based on the European RDS Standard which was adopted in 1982.

Revised editions of both the European RDS Standard, and the U.S. RBDS Standard, were adopted in 1998. To obtain a copy of the U.S. RBDS Standard contact either the NAB or CEMA for ordering information. The European RDS Standard is formally issued by the European Committee for Electrotechnical Standardization (CENELEC) and is issued under Document Number EN 50067.

In the U.S. all references to RDS/RBDS technology generally use the term RDS rather than RBDS. This is because products can be marked with a special RDS logo. RBDS is generally used only to refer to the actual NRSC Standards document. The NRSC felt that it was important to distinguish the U.S. Standard from its European counterpart by giving it a slightly different name. The U.S. standard includes a few features that the European standard does not. One of these is a specification for transmitting Emergency Alert System information via the RDS signal.

The RDS Subcarrier

The FCC has deregulated subcarrier usage allowing almost any technology to be used as long as certain criteria are met. This means that analog or digital or a mixture of both may exist within the subcarrier frequency range as long as the total modulation does not exceed 110%. RDS has proven to be compatible with existing services due to its low injection requirements and narrow bandwidth. In addition, by phase locking the AM double sideband suppressed carrier signal with the stereo pilot signal, there is little risk of interference to the main audio program, even under multipath conditions. RDS is carried on many classical networks throughout Europe and the U.S. with no degradation to the audio quality. Figure 4.6-1 depicts the baseband audio spectrum of a FM broadcast station. RDS nests into the 57 kHz position between the stereo multiplex and the 67 and 92 kHz subcarrier channels. Injection levels as low as 1.3 kHz allow RDS to be easily implemented without giving up program audio power. Since RDS lies at 57 kHz it is compatible with nearly every subcarrier in use or development today. Due to the low implementation cost of RDS, it is not likely to be displaced by another system for many years.

RDS Data Structure

The RDS signal is modulated and demodulated as an asynchronous bit stream. The RDS baud rate is

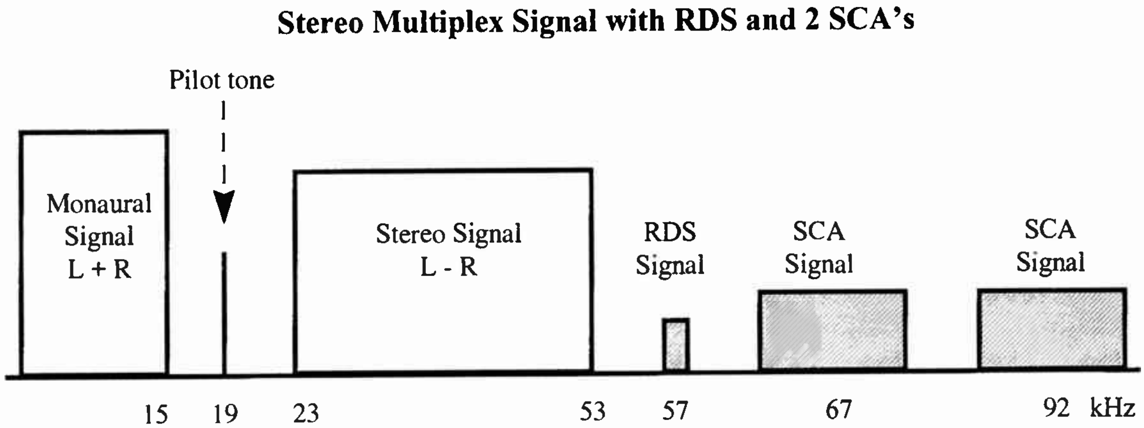


Figure 4.6-1. A typical broadcast system.

1187.5 bits per second. The data transmission is fully asynchronous and there are no gaps between the groups or blocks. The RDS group structure or baseband coding is depicted in Figure 4.6-2. This baseband structure is described as follows:

- The largest element in the structure is called a group (104 bits).
- Each group comprises 4 blocks of 26 bits each.
- Each block comprises an information word and a checkword.
- Each information word comprises 16 bits.
- Each checkword comprises 10 bits.
- All information words, checkwords, binary numbers or binary address values have their most significant bit (msb) transmitted first.

Information Word

The *information word* contains the actual data for each block. This information word is comprised of 16 bits. The most significant bit (m_{15}) is transmitted first.

Checkword and Offset Word

The *checkword and offset word* consist of 10 bits. The most significant bit (c_9) is transmitted first. The checkword allows the receiver to detect and correct errors that occur during data reception. The error-protecting code has the following error-checking capabilities:

- Detects all single and double bit errors in a block.
- Detects any single error burst spanning 10 bits or less.
- Detects about 99.8% of bursts spanning 11 bits and about 99.9% of all longer bursts.
- The code is also an optimal burst error *correcting* code and is capable of *correcting* any single burst of span 5 bits or less.

Figure 4.6-3 shows the possible offset words. The offset word allows detection of the block number within the data group. This allows critical data such as the PI, PTY and TP codes to be decoded without reference to

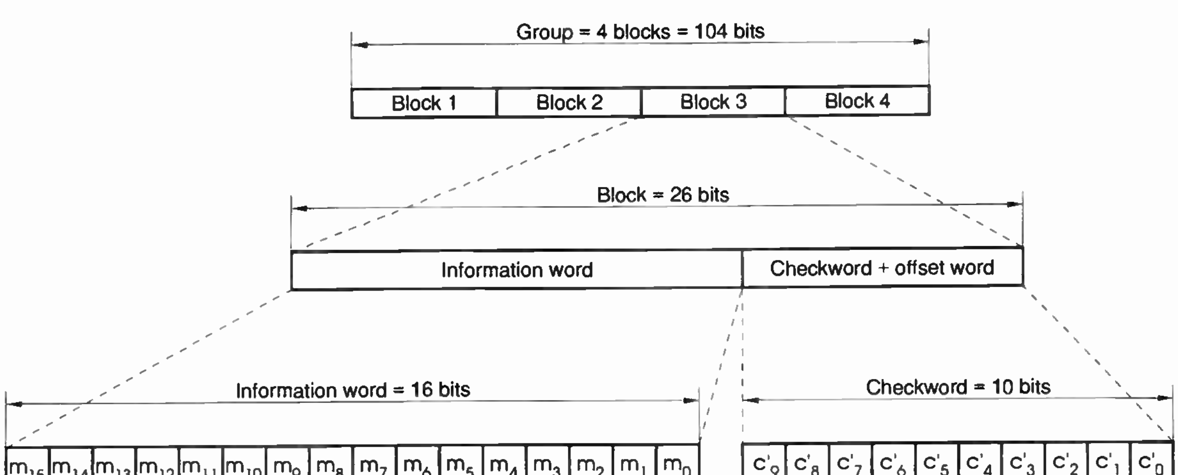


Figure 4.6-2. Structure of the baseband coding.

Offset Word	Binary value d_9	d_8	d_7	d_6	d_5	d_4	d_3	d_2	d_1	d_0
A	0	0	1	1	1	1	1	1	0	0
B	0	1	1	0	0	1	1	0	0	0
C	0	1	0	1	1	0	1	0	0	0
C	1	1	0	1	0	1	0	0	0	0
D	0	1	1	0	1	1	0	1	0	0
E	0	0	0	0	0	0	0	0	0	0

Figure 4.6-3. Offset words.

any block outside the one that contains the information. This is essential to minimize acquisition time for these kinds of messages and to retain the advantages of the short (26-bit) block length. To permit this to be done for the PI codes in block 3 of version B groups, a special offset word (C′) is used in block 3 of version B groups. The occurrence of offset C′ in block 3 of any group can then be used to indicate directly that block 3 is a PI code, without any reference to the value of B_0 in block 2. Offset word E is only transmitted by Mobile Broadcast System (MBS) or RDS/MBS time multiplexing (MMBS) stations. This allows receivers to maintain synchronization in MMBS applications.

Synchronization of Blocks and Groups

The blocks within each group are identified by their offset words A, B, C or C′ and D. The beginning and end of the data blocks may be recognized in the receiver decoder by using the fact that the error-checking decoder will, with a high level of confidence, detect block synchronization slip as well as additive errors. This system of block synchronization is made reliable by the addition of the offset words (which also serve to identify the blocks within the group). These offset words destroy the cyclic property of the basic code so that in the modified code, cyclic shifts of codewords do not give rise to other codewords. A detailed explanation of a technique for extracting the block synchronization information at the receiver is given in Annex C of the U.S. RBDS Standard.

Group Structure

The information coded into each group has a common fixed structure as depicted in Figure 4.6-4.

All data groups share a common fixed structure of information coding to allow critical data to be transmitted in a fixed, highly repetitive pattern to ensure the best reception even under adverse reception conditions. This fixed structure is described as follows:

Block 1—Block 1 of every RDS data group contains only the program identification or PI code.

a) Block 1 is identified by offset word A.

Block 2—Block 2 contains the group type code, version code, traffic program code, and program type (PTY) code. All codes are binary coded with the msb transmitted first.

a) Group type code—The data groups are identified through the 4 bit group type code identified as A3-A0, along with the 1 bit version code, B0. The bits A3-A0 yield decimal values 0 through 15.
b) Version code—Each group then has an "A" version and a "B" version based upon the state of B0 where:
   i) B0 = 0 is identified as group type A
   ii) B0 = 1 is identified as group type B
   iii) All B type groups repeat the program identification (PI) code in blocks 1 and 3.
c) Traffic Program (TP) code—This single bit code is utilized with the traffic announcement (TA) code to provide the traffic feature.

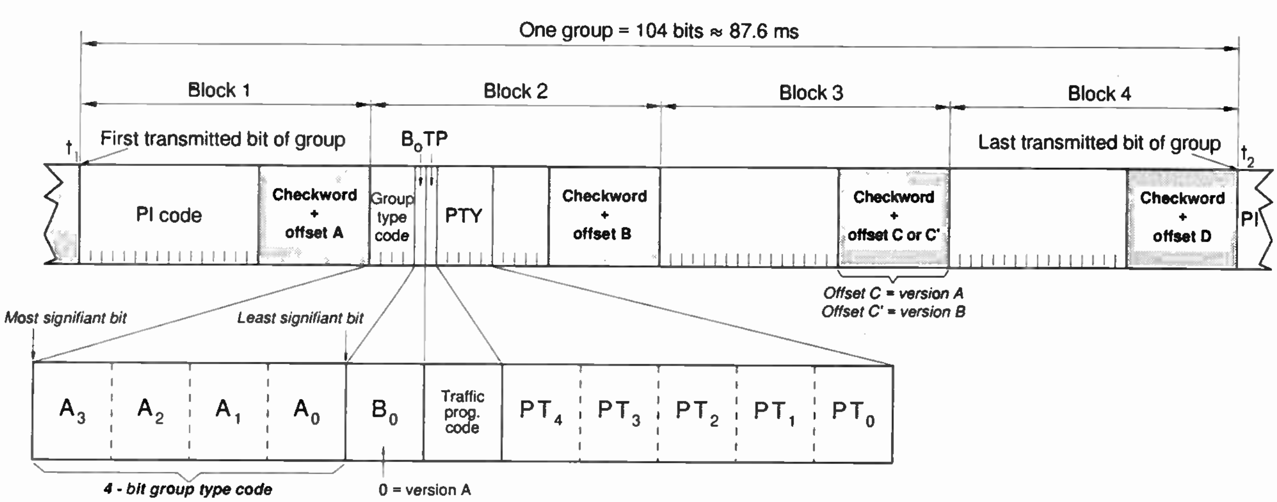


Figure 4.6-4. RDS group coding structure.

d) Program type (PTY) code—This five bit code describes the current audio program being aired. There are 31 possible PTY's available.
e) Block 2 is identified by offset word B.

Block 3—Dependent upon the group version code (A or B) or offset word (C or C'), block 3 may carry two distinct types of data.

a) Version code A—This block carries data defined by the group type code.

Block 3 is identified by offset word C for version code A groups.

b) Version B type groups carry only the PI code.

Block 3 is identified by offset word C' for version code B groups.

Block 4—This block carries data defined by the group type code.

a) Block 4 is identified by offset word D.

RDS Group Types

RDS data groups are referred to by both the group type and version, (0A, 14B). Taking into account both the group type and version codes, there are 32 defined data groups. Some of the group features are defined in their entirety and cannot be defined in any other way, while other data groups are openly defined such that the actual data contained in the group can be defined by the operator. The group type code defines the basic function of the data group. The "A" and "B" versions differ only in the coding of the data for the group. Because RDS data is transmitted in functional groups that share a common structure, the operator can select from these groups types and transmit all or only a portion of them depending upon the features in use. In this manner, there is no need to transmit data which is not in use, thereby maximizing data throughput. Figure 4.6-5 summarizes the possible applications for all possible group types. Only group type 0A or 0B must be included in the transmission sequence.

RDS FEATURES

RDS involves the use of primary and secondary features. Primary features are either contained in every information group or are included in type OA or OB groups, which are always transmitted. Secondary features are transmitted only when necessary or desired. This prevents data capacity from being wasted on unused features. The following is a summary of RDS features.

Primary Features

Program Identification (PI) Code The PI code is a four digit hexadecimal code that is unique for each station. The PI code is calculated from the station's call

Group Type	Group type code/version A_3	A_2	A_1	A_0	B_0	Flagged in type 1A groups	Description
0A	0	0	0	0	0		Basic tuning and switching information only
0B	0	0	0	0	1		Basic tuning and switching information only
1A	0	0	0	1	0		Program Item Number and slow labeling codes only
1B	0	0	0	1	1		Program Item Number
2A	0	0	1	0	0		Radiotext only
2B	0	0	1	0	1		Radiotext only
3A	0	0	1	1	0		Applications Identification for ODA
3B	0	0	1	1	1		Open Data Applications
4A	0	1	0	0	0		Clock-time and date only
4B	0	1	0	0	1		Open Data Applications
5A	0	1	0	1	0		Transparent Data Channels or ODA
5B	0	1	0	1	1		Transparent Data Channels or ODA
6A	0	1	1	0	0		In House applications or ODA
6B	0	1	1	0	1		In House applications or ODA
7A	0	1	1	1	0	Y	Radio Paging or ODA
7B	0	1	1	1	1		Open Data Applications
8A	1	0	0	0	0	Y	Traffic Message Channel or ODA
8B	1	0	0	0	1		Open Data Applications
9A	1	0	0	1	0	Y	Emergency Warning System or ODA
9B	1	0	0	1	1		Open Data Applications
10A	1	0	1	0	0		Program Type Name
10B	1	0	1	0	1		Open Data Applications
11A	1	0	1	1	0		Open Data Applications
11B	1	0	1	1	1		Open Data Applications
12A	1	1	0	0	0		Open Data Applications
12B	1	1	0	0	1		Open Data Applications
13A	1	1	0	1	0	Y	Enhanced Radio Paging or ODA
13B	1	1	0	1	1		Open Data Applications
14A	1	1	1	0	0		Enhanced Other Networks information only
14B	1	1	1	0	1		Enhanced Other Networks information only
15A	1	1	1	1	0		Do not use; Fast PS being phased out.
15B	1	1	1	1	1		Fast switching information only

Figure 4.6-5. RDS group types.

letters, thus no two are alike. The receiver uses the PI code to identify a station rather than frequency. RDS allows frequency diversity such that if a station simulcasts on another frequency, the receiver can automatically tune to the strongest station. If a station simulcasts, then it must pick one station's PI code and use it on all of the simulcasting stations. While the PI code is transparent to the user, its proper use is vital to proper receiver operation. Specialized PI codes are available for use for networked programming as well.

Program Service (PS) Name The PS Name is whatever name the station chooses to present to its listeners. It can be call letters like "WZWZ-FM," or a slogan like "Z-93." The PS Name is displayed instead of frequency on an RDS receiver. When the listener tunes to an RDS station the station's name will be shown on the radio display. The PS Name cannot be longer than eight characters. Any character referenced in the RBDS Standard may be used, but special characters might not be displayable on certain low cost displays. The note appearing at the bottom of Table E.1 of the RBDS Standard details which characters are most commonly displayable.

Traffic Program (TP) If a station offers listeners traffic bulletins then it should pay particular attention to this feature. The TP identifies the transmitting station to the listener as a station that offers traffic programs. RDS receivers can automatically tune to stations that offer traffic programs when the user turns on the traffic announcement feature of the radio. A broadcaster should not set TP high unless it actually uses the TA feature.

Traffic Announcement (TA) When an actual traffic bulletin is broadcast this information bit must be set to a logic "1." The RDS receiver detects this and will automatically stop any playback device that may be in use, and will return to the FM tuner mode. Audio adjustments are also automatically made so that, if the user had the volume muted for instance, the receiver will adjust the volume to a user preset level. Some means must be employed to control the state of this information bit coincident with the beginning and ending of the actual traffic announcement. Figure 4.6-6 depicts the proper usage of the TP and TA bits during a traffic announcement.

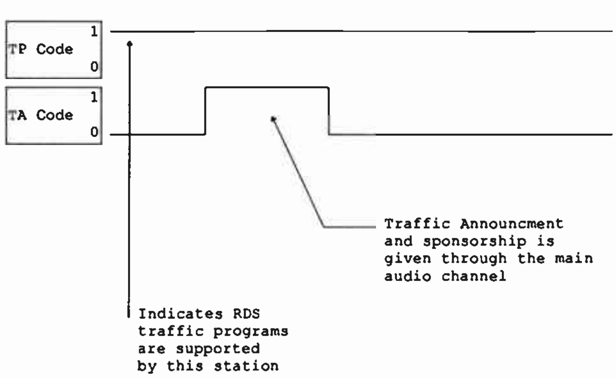


Figure 4.6-6. A typical non-EON traffic announcement.

Program Type Codes (PTY) This code is used to identify the type of program material being broadcast. There are pre-defined codes for *Country, Rock, Top 40* etc. The RDS receiver can automatically search for stations by PTY, thus allowing listeners to find their favorite type of programming without tuning to all available stations. Advanced receivers can even interrupt listeners when, for example the *News* is broadcast. A definition of the terms used to denote North American Program Types is given in Table 4.6-1.

Alternate Frequencies (AF) This feature allows the RDS receiver to automatically tune to the best signal when multiple transmitters or translators are broadcasting the same program. Regional or national programs that are broadcast over large areas can even be linked together, providing the listener with the illusion of one very powerful transmitter. It is also possible to only link stations during specific times when the program material is common (perhaps during a sporting event) using the linkage actuation feature. Competitors cannot steal listeners away using this feature because an RDS receiver will only look for alternative frequencies that are specified by the station that is currently tuned in. An example of an AF network is shown in Figure 4.6-7.

Secondary Features

Program Type Name (PTYN) This feature allows the ultimate in flexibility for the broadcaster who wants to stand out from competitors. While the PTY codes are predefined, the PTYN can be any eight characters the broadcaster desires to further describe the current program. For example, a broadcaster who is currently using the PTY "Personality" may set the PTYN to "Limbaugh" to further describe the current program. A "Rock" station may set the PTYN to "Bob & Tom" to describe their morning team. An RDS receiver cannot search by PTYN, but will display the PTYN in place of the PTY once tuned to a particular station.

Radiotext (RT) Radiotext allows a station to transmit up to 64 characters of information to the listener. Information such as the current artist and song title, station promotional information, local events, and even additional information about the advertiser whose commercial is currently playing can be sent.

Clock Time and Date This feature enables broadcasters to transmit the current date and time once every minute, thus enabling RDS equipped clock radios, car radios, etc., to be automatically set after a power failure, time zone change, etc. Broadcasters should not use this feature unless they can ensure the accuracy of the transmitted information within ± 2 sec.

Emergency Warning System (EWS) The EWS feature allows the transmission of coded emergency information intended for specialized receivers. The FCC has recognized the capability of this feature, and permits broadcasters to use RDS to transmit Emergency Alert System information. Broadcast EAS equipment can be linked to an RDS encoder for automatic retransmission of EAS information. (However, this does not eliminate the requirement to transmit EAS information in the main program channel as well.)

Table 4.6-1

Program Type	Name	Description
1	News	News reports, either local or network in origin.
2	Information	Programming that is intended to impart advice.
3	Sports	Sports reporting, commentary, and/or live event coverage, either local network in origin.
4	Talk	Call-in and/or interview talk shows either local or national in origin.
5	Rock	Album cuts.
6	Classic Rock	Rock oriented oldies, often mixed with hit oldies, from a decade or more ago.
7	Adult Hits	An up-tempo contemporary hits format with no hard rock and no rap.
8	Soft Rock	Album cuts with a generally soft tempo.
9	Top 40	Current hits, often encompassing a variety of rock styles.
10	Country	Country music, including contemporary and traditional styles.
11	Oldies	Popular music, usually rock, with 80% or greater non-current music.
12	Soft	A cross between adult hits and classical, primarily non-current soft rock originals
13	Nosotalgia	Big band music.
14	Jazz	Mostly instrumental, includes both traditional jazz and more modern smooth jazz.
15	Classical	Mostly instrumentals, usually orchestral or symphonic music.
16	Rhythm And Blues	A wide range of musical styles, often called *urban contemporary*.
17	Soft Rhythm	Rhythm and blues with a generally soft tempo.
18	Foreign Language	Any programming format in a language other than English.
19	Religious Music	Music programming with religious lyrics.
20	Religious Talk	Call-in shows, interview programs, etc. with a religious theme.
21	Personality	A radio show where the on-air personality is the main attraction.
22	Public	Programming that is supported by listeners and/or corporate sponsors instead of advertising.
23	College	Programming produced by a college or university radio station.
24-28		Unassigned.
29	Weather	Weather forecasts or bulletins that are non-emergency in nature.
30	Emergency Test	Broadcast when testing emergency broadcast equipment or receivers. Not intended for searching or dynamic switching for consumer receivers.
31	Emergency	Emergency announcement made under exception circumstances to give warning of events causing danger of a general nature. Not to be used for searching—only used in a receiver for dynamic switching.

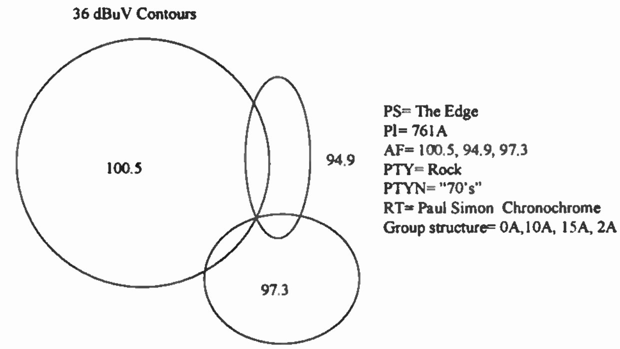


Figure 4.6-7. A typical RDS alternate frequency network.

An alert feature contained within consumer receivers will be activated when an alert code is transmitted. The U.S. RBDS Standard includes an entire annex devoted to implementing EAS within RDS.

Open Data Channel (ODC) This is the latest and most flexible addition to the RDS Standard. In previous years working groups were formed to determine the best way to add additional data services such as *differential correction* for the Global Positioning System (DGPS), and even the EAS. These systems needed both public (free) and private (fee) based delivery. Using the remaining few undefined RDS data groups to perform these services proved too much, as coordinated systems could not be agreed upon. However, it was agreed that each data group left undefined was worth its weight in gold, since once all the groups were defined, that meant the end of any future expansion to the system. Thus was born the concept of the open data channel.

The open data channel allows the definition of an unused RDS data group based solely on the Application Identification (AID) code. Thousands of AID codes are available meaning that the remaining data groups could be defined thousands of different ways. The AID codes are internationally assigned and coordinated so that anyone can apply for an AID code and start their very own data service. The ODC is much more flexible than other data only groups such as the transparent data channel because it can be automatically tuned (by use of the AID) and tracked regardless of any other data being transmitted. This is a dynamic, and potentially very profitable, data group.

Transparent Data Channel (TDC) For specialized applications, information of any type can be transmitted by use of this data group. For instance, advertising messages could be sent to an electronic billboard using this feature. Virtually any way that data is sold can be supported by the TDC. Normal car and home receivers do not decode this information.

In-House Application Data contained in this group is to be used only by the broadcaster. Remote control applications, station telemetry, or paging applications can be supported using this data group.

Radio Paging (RP) Paging services including numeric or alphanumeric can be supported by RDS.

Traffic Message Channel (TMC) This feature allows traffic information to be coded and displayed on consumer receivers so that normal audio bulletins are not required

SUMMARY OF CHANGES IN THE NEW STANDARD

The newly revised European RDS and U.S. RBDS Standards move forward into a new era for FM data services. All changes have been devised to be fully backwards compatible with the old standard while offering a tremendous amount of growth. The prior standards were:

- United States RBDS Standard—January 8, 1993. Issued by the National Radio Systems Committee (NRSC)
- EN50067, Specification of the radio data system (RDS)—April 1992. Issued by the European Committee for Electrotechnical Standardization (CENELEC)

While these standards shared much commonality, the newly revised standards are almost completely identical in form. To utilize the new features contained in the 1998 version of the U.S. RBDS Standard, a software upgrade must be made to existing broadcast encoders. It is a good idea to inquire about upgrades before making a purchase of new equipment as well. The following features were modified in the new RBDS Standard:

- Program Type Name (PTYN)
- Fast Program Service Name (Fast PS)
- Location and Navigation (LN)
- Open Data Applications (ODA)
- Enhanced Radiopaging Protocol (EPP)
- Language Identification (LI)
- Decoder Information (DI)
- Extended Country Code (ECC)
- PI structure from B000-FFFF
- Analog SCA cross referencing
- IRDS Updating
- PTY code table
- MBS EAS

Program Type Name (PTYN)

The PTYN contained within group 10A has been added to the RDS standard. This feature was previously defined only within the RBDS Standard. The feature has been slightly modified with the addition of a text A/B Flag, but is still backwards compatible with the previous definition. The A/B flag is located in the unused 4 bit region and shares the same bit position as the A/B flag in Radiotext group 2A. The usage is the same in both groups in that the A/B flag should be toggled whenever the PTYN is changed. Within the receiver, the PTYN text buffer should be cleared when the A/B flag is toggled to prevent *mixing* of old and new PTYN's. The rationale is that the PTYN is designed to be semi-dynamic, potentially changing whenever the program changes from "news" to "weather" to "sports"). Newly designed receivers can take advantage of the A/B flag without effects from existing encoders that do not incorporate this change.

Fast Program Service (PS) Name

The Fast PS is currently defined in group type 15A of the RBDS Standard and is not defined in the RDS Standard. This group will not be adopted in Europe. In North America there will be a phase out of all transmissions within 10 years, and 2 years for new encoders. The elimination of this group can be included in encoder upgrades to the new standard. This group will not be defined for any use and could later be reallocated as open data. This will not adversely affect the existing receiver base as only Delco Electronics, Grundig, Blaupunkt, and Becker have produced receivers for North America with this feature. The rationale for this change is that it will not cause significant performance degradation in the receivers as the PS is also located in groups 0A, and 0B. The PS is transmitted two characters at a time in these groups and will be received in 348 ms versus 174 ms using the fast PS feature. Since the PS is typically static and is often stored in receiver memory for instant recall following PI reception, the only time that slower PS acquisition will be noticed is when tuning to a station for the first time. There are over 35 million receivers in use worldwide that would not be able to receive the fast PS. Once broadcasters begin using Fast PS, acquisition times of PS in existing receivers would be adversely impacted since they would not be able to decode this data. Therefore as a feature, Fast PS cannot be considered fully backwards compatible. The Fast PS would be of minor benefit especially in Europe where most broadcasts are network based and hence require frequent repetition of AF information (0A). So in consideration for phase out in North America, the Europeans have agreed to coordinate the future use of 15A.

Location and Navigation (LN)

LN was defined in Group 3A in the RBDS Standard only. This usage will be immediately deleted and reassigned to the Open Data Application (ODA). This was recommended since there were no LN transmissions within North America and this feature was no longer required for ID Logic receivers. The definition as LN was found to be inadequate by the Differential Global Positioning System (DGPS) Working Group of the NRSC. This group along with its European counterpart determined that the LN structure could not meet current needs of private and public service providers. It was recommended that the ODA be used for LN and DGPS services.

Open Data Applications (ODA)

Group 3A was redefined for ODA. Other, previously unused groups were assigned for ODAs as well. The use of an "A" type group allows low data capacity application to exist entirely within Group 3A. The ODA allows multiple reuse of unused data groups rather than fixed definitions, ensuring a long future for RDS. The actual use of data can be either public or private. The ODA allows encryption of data for fee based services. Data groups referenced via slow labeling (Group 1A) can also be used for ODA when not specifically referenced for use. Specific uses are identified by application identification codes (AID). Over 65,000 internationally allocated AID Codes are available. This allows anyone to start their own datacasting business locally, nationally, or internationally. ODA are already defined or proposed for EAS, DGPS, DAB cross referencing and enhanced TMC.

Enhanced Radiopaging

To enhance the existing paging capabilities of RDS, the enhanced radiopaging protocol was developed. This protocol is contained within Groups 13A and 1A. Annex M of the U.S. RBDS Standard contains the protocol definitions. With the enhanced protocol the possibility of international RDS paging has been added. The prior definition of RDS paging was based upon the usage of PI codes only. This worked well in network based broadcasts as found in many parts of Europe, but offered no solution where non-network based broadcasts are employed, such as in the U.S. The enhanced radiopaging protocol now opens local, national and international paging opportunities.

Language Identification

To aid the international traveler, the language identification feature located in Group 1A Variant 3 has been expanded to include most world languages. The list of covered languages is located in Annex J of the Standard. The feature enables the broadcaster to indicate the spoken language being broadcast thus allowing listeners to search for their native language on local broadcasts.

Decoder Identification (DI)

The decoder information feature has been modified to enable the identification of stations that perform dynamic PTY switching. This allows the broadcaster to indicate that it dynamically switches the program type (PTY) based on the current audio program thus supporting the use of PTY interrupts in consumer receivers. Prior to this, there was no way to tell if the broadcaster would ever change the PTY. The dynamic PTY indicator is analogous in use to the TP bit of the traffic announcement feature.

Extended Country Codes (ECC)

ECC allow receivers to identify the country that the broadcast is coming from. Since PI codes are limited in number, they must be repeated throughout the world, when the PI code is received in conjunction with the ECC, the exact country of origin can be identified. The updated ECC code table has been expanded to be international in scope. It is contained in Annex D and Annex N of the Standard. ECC should be transmitted by all broadcasters, and it is recommended that it be a default automatic transmission in encoders.

Program Identification (PI) Codes

In the U.S. PI codes are based on call letters rather than being assigned by any organization as is done in Europe. A portion of the PI codes are reserved for network usage and also for assignment to stations in Canada and Mexico. During the upgrade to the U.S. RBDS Standard, several mistakes were discovered in the non-call based PI codes which had to be corrected. The changes to the PI code assignment are summarized as follows:

- PI assignments below B000 will remain as is, allowing AF switching but no regionalization. PI codes "_0__", and "__00" are re-mapped into the "A" range of PI's.
- C000-CFFF assigned to Canada. Allows AF switching, but no regionalization. PI codes C0xx, and Cx00 are excluded from use.
- F000-FFFF assigned to Mexico. Allows AF switching, but no regionalization. PI codes F0xx , and Fx00 are excluded from use.
- B_FF, D_01-E_FF assigned for national networks in U.S. and Canada, Mexico. Regionalization allowed. NRSC to provide assignments for all three countries.

The ECC code table was modified for these changes as well. The rationale for these changes was that the current definitions for PI's above BFFF were contradictory. Also, the present PI code structure above BFFF were inadequate for the needs of Canada and Mexico because the existing PI structure only allowed 256 PI's for Canada and Mexico. The new PI assignments yield 3,584 possible non-regional PI's for Canada and Mexico. As well as 765 national network PI"s for all three countries.

Analog SCA Cross Referencing

This feature was devised to allow RDS signaling to receivers equipped with analog subcarrier decoders to switch over automatically. This feature was defined in section 3.1.3.6, Note 3 as part of the transparent data channel (TDC), Group 5A Channel 2. This feature has been deleted from the revised standard because: a) it is not in use; b) it prevents harmonization with the European RDS Standard; and, c) it can be easily converted over to the ODA channel.

ID Logic RDS (IRDS) Updating

This feature was previously defined in section 3.1.3.6 Notes 1 and 2 as a reservation of the transparent data channel (TDC), Group 5A Channel 0 and 1. This was reserved for updating ID Logic receiver databases through RDS. The TDC reference was deleted from the revised standard. The rationale behind this change was that it was not currently in use and it prevented harmonization with the European RDS Standard. ODA will now be used for updating ID Logic receiver databases. The use of AID codes will provide better protection and international usage of the IRDS feature.

PTY Code Table

Both the RDS and RBDS Standards have included additional definitions for previously undefined PTY codes. RBDS has defined two new codes:

- PTY-29, Weather. It is defined for non-emergency weather related information such as forecasts, watches, and advisories. Weather is a popular listener feature that can be supported with the PTY "Watch" mode to interrupt playback similar to traffic announcements and News PTY.
- PTY-23, College. Is intended to be used to identify broadcasts oriented to college students that often contain a variety of formats and information. It is one of the top three non-commercial formats. The other two, Public and Religious, already had PTY definitions.

In the European RDS Standard all the PTY codes have now been defined. The main reason for this was to ensure coordination with the development of the digital audio broadcasting system being rolled out throughout Europe known as the Eureka 147 Standard. Previously, codes 16 through 30 had no definition. These codes have all been defined in the new standard.

DIFFERENCES BETWEEN RDS AND RBDS

With all the changes to the two standards, there still remain differences. These differences are mainly market driven. From a listener point of view, most of these differences are transparent. In exploring these differences it will become evident that the receiver manufacturer must be well versed in the differences to ensure proper receiver operation. When these differences are properly employed by the receiver manufacturer, it is possible to construct a truly global receiver. If broadcasters and receivers make use of the ECC feature, it is possible that the receivers could self-configure automatically. The problem of course, is getting broadcasters to utilize the ECC feature since it is not a required broadcast feature.

Summary of Differences Between European RDS and U.S. RBDS

- Program Type Definitions (PTY)—To conform with differing broadcast styles, the PTY code definitions are different. These differences may be accounted for through the use of a look-up table within the receiver. Annex F of the U.S. RBDS Standard includes both the European and North American PTY definitions.
- Program Identification Coding (PI)—Due to the high penetration of network based programming throughout Europe as compared to the largely independent single transmitter structure of North America, the derivation of PI codes is different. In the RBDS Standard, PI codes are based on call letters from 0000-AFFF, and are network based from B000-FFFF. This means that PI codes above B000 in North America are treated by the receiver the same way that all PI codes are treated in Europe. Below B000, PI codes in North America do not employ the regionalization feature. While alternate frequency switching is still employed below B000, variants based upon changes in the second nibble of the PI code, do not exist, and should be ignored. PI code definitions can be found in Annex D of the Standard.
- Mobile broadcast system (MBS) and RDS/MBS time multiplexing (MMBS)—The predecessor to RDS, MBS, still thrives throughout North America. MBS is mainly used as a paging system through a network of approximately 500 stations within the U.S. The MBS system utilizes the same modulation and data structure as RDS, but employs a different data protocol. An MBS broadcast is identified through the offset word E. Since there are similarities between the two systems, it is possible to time multiplex MBS and RDS data. This time sharing is known as MMBS. Receiver manufacturers must be able to differentiate between RDS and MBS, as well as accommodate MMBS broadcasts. Internal MBS/MMBS cross references can be found throughout the new RBDS Standard as a reminder of particular system requirements and as a possible alternative to RDS. A public domain EAS protocol is also contained within the MBS/MMBS annex.
- EAS ODA Protocol—Within the U.S. RBDS Standard, the EAS ODA protocol is defined for use in the U.S. This optional feature set is constructed around the Federal Communication Commission's EAS protocol and is open for public use. RDS allows the silent retransmission of emergency information. This has been combined with existing consumer oriented emergency features to allow additional feature functionality to consumer receivers as well. The EAS ODA can also accommodate private emergency systems.
- ID Logic (IDL) in — Receiver database updates via RDS (IRDS) ODA Protocol—The ID logic feature is a licensed technology that allows the incorporation of an in receiver database that contains format type and call letters for all AM and FM stations. When combined with RDS, IDL can provided similar data and features for non-RDS and AM stations. The IRDS feature allows the database to be updated through an RDS ODA so that the information can be updated and maintained automatically. The introduction of ID Logic is contained in Section 7 of the U.S. RBDS Standard, and the IRDS ODA is described in detail in Annex R.
- AM RDS (Future System)—Although a suitable data transmission system has yet to be developed that is compatible with the C QUAM AM stereo system, there is a separate section in the U.S. RBDS Standard where such a system can be defined or referenced. The NRSC remains interested in the development of an AM RDS system.

BIBLIOGRAPHY

This chapter was adapted from the book *The Broadcasters Guide to RDS,* Scott Wright, Focal Press, 1997, ISDN 0 240 802780.

Specification of the Radio Data System (RDS) for VHF/FM Sound Broadcasting in the Frequency Range 87.5 to 108.0 MHz, EN 50067:1998, European Committee for Electrotechnical Standardization (CENELEC). Brussels, Belgium.

Specification of the Radio Data System, EN 50067:1992, *European Committee for Electrotechnical Standardization* (CENELEC), Brussels, Belgium.

United States Standard, April 9, 1998, National Radio Systems Committee.

United States RBDS Standard, January 8,1993, National Radio Systems Committee.

4.7
TRANSMISSION SYSTEM CONTROL AND MONITORING

HAROLD HALLIKAINEN
HALLIKAINEN & FRIENDS, INC., SAN LUIS OBISPO, CA

INTRODUCTION

The Federal Communications Commission (FCC) requires broadcast transmitters to have various meters and controls. The required meters include those necessary to determine the transmitter power output (or input to the antenna system) by the direct and/or indirect methods. These are normally those instruments necessary to determine the transmitter final amplifier input power (final amplifier voltage and current) and the output power (transmission line meters for FM and TV, base or common point meters for AM).

On some transmitters, measurement of the final amplifier voltage and current is meaningless. In these cases, the transmitter manual should describe what the FCC required for acceptance of the transmitter.

The required controls include those necessary to insure that the transmitter is operating within the terms of the station license. These controls have traditionally included a carrier on/off control, a day/night control for AM stations (power and/or pattern change) and a *power trim* control. Directional AM stations and television stations must have additional monitors. Directional AM stations are required to have an FCC authorized antenna monitor (Section 73.69). This monitor indicates the current ratios and phases for each tower, relative to a reference tower. Television stations are required to have a monitor capable of determining whether the transmitted visual signal meets FCC specifications (Section 73.691).

In 1995, the FCC made major changes to its requirements for transmitter control. (see *Report and Order*, MM Docket 94-130, 60 FR 55476). Section 73.1400 was rewritten and renamed *Transmission system monitoring and control*. The revised rules allow stations to be operated in either an *attended* or *unattended* mode.

Stations operating attended have a *designated person* in charge of the transmitter. There is no longer a requirement that the person controlling the transmitter have an FCC operator's license. The person in charge of the transmitter may observe and control it directly (if the transmitter is nearby) or by using a remote control. Finally, the person in charge may merely supervise an *automatic transmission system* (ATS) (Section 73.1400(a)(2)). The ATS is required to monitor and control (licensee determined) critical parameters that would cause interference should they be out of tolerance. Should an interference causing condition exist that the ATS cannot correct, the ATS is required to notify the person in charge of the transmitter. The previous requirement that an ATS shut down a transmitter has been removed for stations operating in an attended mode with an ATS.

Stations operating unattended rely upon stable equipment or an ATS instead of relying upon a person to insure the transmitter operates within limits. If an ATS is used, it monitors and controls licensee determined critical parameters. Should an interference causing condition exist, the ATS shuts down the station within three hours. ATS systems at unattended stations are not required to notify anyone before shutting the transmitter down. However, most stations want someone to be notified if the transmitter is going to be shut down, so they set up their ATS systems to do this anyway.

FCC rules previously authorized *extension metering*. Extension metering was a form of remote telemetry where a dedicated circuit was used for each parameter being monitored. No control functions were required to be extended. Instead, a distance limit on extension metering permitted the transmitter controls to be reached quickly. The adoption of the rules permitting unattended operation deleted the extension metering provisions.

This chapter discusses the sampling and control of various transmitter parameters. These sampling and control concepts can be applied to remote control (including *extension metering*) and ATS. The discussion then moves on to data conversion and transmission subjects important to remote control and remote ATS. This is followed by a discussion of the operator interface, and data analysis. Although the vast majority of transmitter control systems available today are microprocessor based, nonprocessor based systems are presented to develop the concepts common to all systems and to provide an historical perspective. The chapter organization can be thought of as starting at a transmitter and working back to the control point.

PARAMETER SAMPLING CIRCUITRY

Most transmitter parameters can be reduced to a *sample voltage* by using voltage dividers or current sense resistors. *Remote metering* is accomplished by sending a signal representing this sample voltage to the remote metering point, then displaying a value representing the original parameter. The concept of sample voltages is important. Although the parameter could be measured directly from the remote metering point, wiring costs often make this impractical. For example, it would be impractical to send 5,000 V to the remote metering point to measure the final amplifier plate voltage of a vacuum tube transmitter. Nor would it be practical to send several amperes of plate current or antenna current to the remote metering point. Instead, voltages (typically dc) representing the values of these parameters are sent. The scale of the remote meter can be calibrated such that every volt of sample voltage represents another kilovolt of plate voltage. The remote meter can directly indicate the transmitter plate voltage without having to deal with high voltages.

In practice, the actual sampling ratio (sample voltage/parameter value) is not important as long as the sampling ratio is stable and the sample voltage is reasonable (low enough to be easily handled, high enough to not get lost in noise). Most transmitter manufacturers provide remote samples for the FCC required indicating instruments. Figure 4.7-1 shows some typical voltage sampling circuits.

DC Voltage Sampling Circuits

Figure 4.7-1(a) shows how a voltage divider may be used to divide the plate voltage down to a safe sample voltage. This circuit is independent of the transmitter front panel plate voltage meter.

In extension meter applications, broadcasters may want to adjust R2 to provide a voltage that is an integer order of magnitude below the actual plate voltage. For example, the extension meter could be a 3.5 digit digital panel meter (a DPM that reads between −1,999 and +1.999), which typically reads in millivolts. R2 is adjusted such that the sample voltage is 300 mV when the plate voltage is 3.00 kV. The appropriate decimal point on the display can be enabled to directly indicate the plate voltage in kilovolts with a resolution of 0.01 kV.

Figure 4.7-1 (b) shows how the existing plate voltmeter *multiplier resistor* (R1) can be used to drive the front panel meter and provide a sample voltage. The additional 1K sampling resistor (R2) decreases the reading of the local meter by 0.01% (which is negligible).

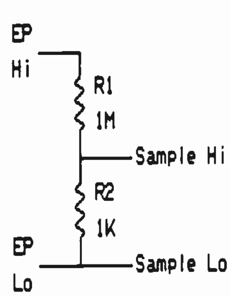


Figure 4.7-1(a). Unbalanced plate voltage sampling circuit. Sample voltage = EP/1001.

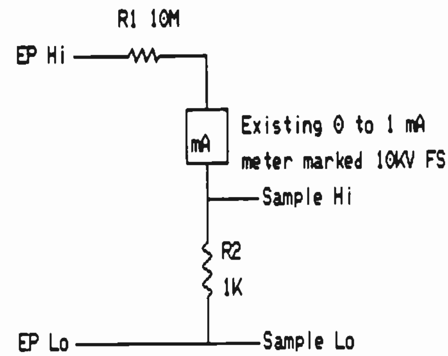


Figure 4.7-1(b). Adding resistor to existing plate voltmeter circuit provides a remote sample.

Clamp Diode for Safety

In each of the sampling circuits shown in Figure 4.7-1, the full plate voltage would be supplied to the remote indicator should the shunt resistor (R2) open (assuming minimal loading by the remote indicator). To protect wiring and personnel, it is common practice to add a *zener diode* across R2. Should R2 open, the sample voltage is clamped to the zener voltage. The zener will typically fail to a short, protecting the external sampling equipment. This zener is commonly damaged in lightning storms, causing a loss of plate voltage sampling.

Common Mode Voltages

Input and output terminals in the circuits of Figure 4.7-1 are labeled "high" and "low." The low terminal is the one whose voltage is closest to ground. In many cases, the low terminal connects directly to ground, allowing the sample voltage to be measured between sample high and ground. Many transmitters, however, do not measure the plate voltage between the plate supply and ground. These transmitters often have resistors between the final amplifier tube cathode (or transistor emitter or source) and ground. These resistors are used to provide protective bias, cathode current sampling and output power trimming. The final amplifier input voltage is then measured between the plate supply to the tube (generally +3 kV to +10 kV) and the cathode of the tube (0 V to +250 V). The circuits of Figure 4.7-1 could possibly place the low side of the sample 250 V above ground. This hazardous voltage would require special precautions in wiring. Further, many remote control systems assume the low side of the sample is grounded. Connecting such a remote control to such a transmitter would short out the circuitry in the cathode circuit, disabling the tube protective bias and current overload protection and causing the transmitter to run maximum power, regardless of the power trim setting.

Common Mode Attenuator

Figure 4.7-2(a) shows circuitry similar to that presented so far. It is driven by +5 kV from the plate of

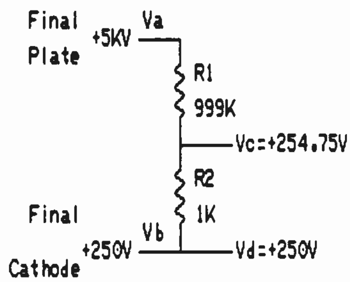


Figure 4.7-2(a). Plate voltage sampling circuit without common mode attenuation gives dangerous sample voltage.

the amplifier and +250 V from the cathode. The *plate voltage* is the differential voltage (Va − Vb). The differential voltage is attenuated (multiplied by 0.001) to arrive at the differential sample voltage. Unfortunately, the circuitry shown does not attenuate the common mode voltage ((Va + Vb)/2) as much, multiplying it by 0.096. This yields the dangerously high common mode voltage.

Figure 4.7-2(b) shows a sampling circuit that attenuates both the common mode and differential mode voltages equally. As in Figure 4.7-2(a), the differential mode sample voltage is 0.001 times the differential mode input voltage (0.001 times 4.75 kV or 4.75 V, in this case). However, the common mode sample voltage is also multiplied by .001, reducing it from 2.625 kV to 2.625 V. This circuit can be used to sample high voltages that are not measured with respect to ground provided the remote metering equipment has differential inputs.

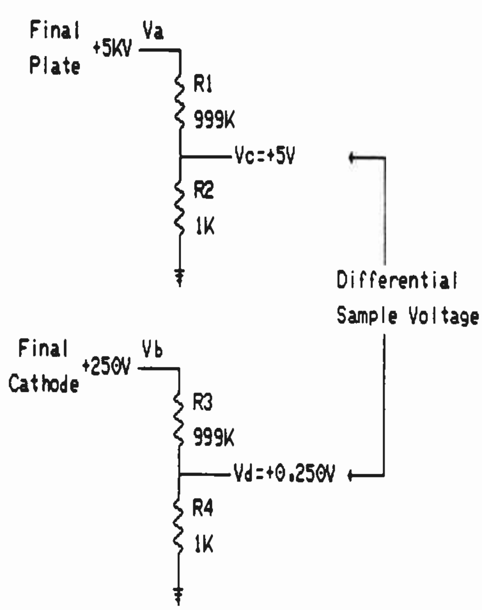


Figure 4.7-2(b). This voltage divider reduces common mode voltage to a safe +2.65 V.

Common Mode Rejection

The differential output voltage of Figure 4.7-2(b) (Vc − Vd) is independent of the common mode input voltage ((Va + Vb)/2) if all the resistors are precisely matched. This can be checked by applying a high common mode voltage with no differential voltage. When both Va and Vb are connected to the high voltage supply (+5 kV), one of the 1K resistors can be adjusted as needed to give no voltage between Vc and Vd.

Differential Input Circuitry

Several techniques are available for dealing with differential sampling voltages (where neither side of the sample is grounded). For circuit simplicity, many remote control systems assume one side of the sample voltage is grounded. Other systems provide differential inputs. If differential samples must be measured, additional circuitry may be added outside the remote control to convert the sample to a ground referenced sample. Typical approaches to dealing with a differential input voltage include: use of an isolation amplifier, use of a differential amplifier, floating the remote control analog to digital converter and calculating the differential voltage.

Isolation Amplifier

An *isolation amplifier* may be included inside a transmitter, a remote control, or between the two units. Isolation amplifiers provide isolation between the input and output by using magnetic, capacitive, optical or thermal coupling between input and output. Many isolation amplifiers require a floating power supply to drive the input circuitry, which considerably adds to cost. Careful design is required to insure that the *encode* and *decode* processes track each other (the output voltage is indeed proportional to the input voltage). Some isolation amplifiers use feedback techniques to reduce the nonlinearities induced by the coupling method.

Differential Amplifier

In most cases, electrical isolation between the sampling circuitry and the remote metering equipment is not required. Instead, a differential voltage must be changed to a *single ended* voltage (one side grounded). This can be accomplished using one of the differential amplifiers shown in Figure 4.7-3. Figure 4.7-3(a) is a basic differential amplifier. If the resistors are precisely matched, the output voltage will be a constant (the differential gain) times the differential input voltage. If the resistors are not precisely matched, the amplifier will have a *common mode gain* other than 0. The output voltage is (Va − Vb)Ad + ((Va + Vb)/2)Acm, where Ad is the differential gain (amplification factor) and Acm is the common mode gain. If the resistors are precisely matched, Acm = 0.

The circuit of Figure 4.7-3(a) suffers from a limited input voltage range. If the amplifier is setup for a gain of 1 (R1 = Rf), the circuit will only accept about twice the input range of the operational amplifier. The circuit of Figure 4.7-3(b) allows up to 200 V of common mode input voltage.

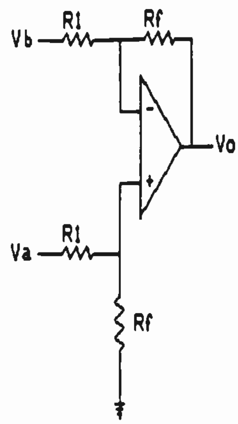


Figure 4.7-3(a). Basic differential amplifier. Vo = (Rf/R1)(Va – Vb).

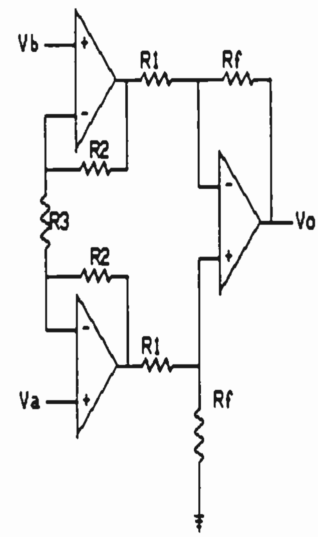


Figure 4.7-3(c). Standard instrumentation amplifier. Gain can be varied by changing R3.

The circuit of Figure 4.7-3(c) (the instrumentation amplifier) loads the sample circuitry less than the other circuits, has better common mode rejection (lower common mode gain), and the gain can be altered by changing only one resistor (R3). The input voltage range is, however, limited to the voltage range of the operational amplifiers.

Calculated Differential Voltage

Each of the schemes so far has been directly measuring the differential voltage or calculating it using operational amplifier (analog computer) techniques. By reconsidering Figure 4.7-2(b) and measuring the sample voltages at points C and D with respect to ground (Vc and Vd), the differential voltage can be calculated. A typical remote control could put Vc on one channel and Vd on another channel. The remote indication would show Vc at 5 V, which indicates Va is 5 kV. The remote indication would show Vd at 0.250 V, which indicates that Vb is 250 V (0.250 kV). By subtracting, the transmitter operator could determine the plate voltage (5.000 kV − 0.250 kV = 4.750 kV). If the remote control has a software subtraction capability, the differential plate voltage could be displayed directly.

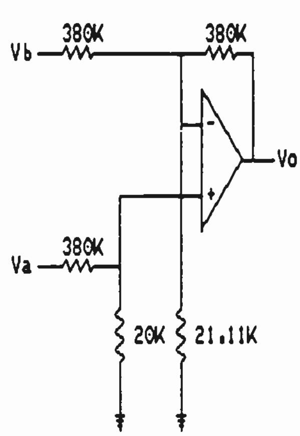


Figure 4.7-3(b). Burr-Brown INA117 differential amplifier with 200 V input range. V0 = Va − Vb.

This approach does require the voltages to be relatively stable, as most systems do not sample all the channels simultaneously.

DC Current Samples

Figure 4.7-4(a) shows the use of a current sense resistor to give a voltage sample that is proportional to current. Unfortunately, the common mode voltage of this sample is very high (about 5 kV). An isolation amplifier would probably be required to safely utilize this sample.

Of course, the plate current also flows through the cathode, so a ground referenced sample is available there. However, the cathode current also includes the control grid current and screen grid current (if the tube has a screen grid). If the grid and screen circuits are returned above the sense resistor, an isolated sample of plate current is available (as shown in Figure 4.7-4(b)). If these circuits are not returned above the sense resistor, samples for each of these currents can also be derived. These currents can then be subtracted from the measured cathode current using analog or software techniques.

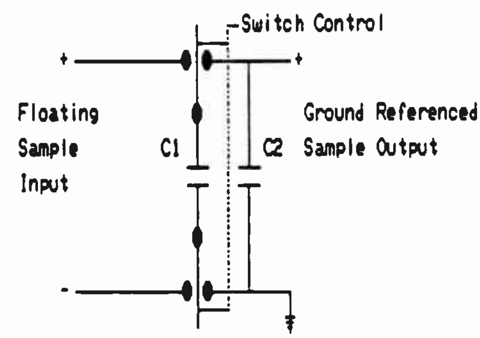


Figure 4.7-3(d). Flying capacitor isolator.

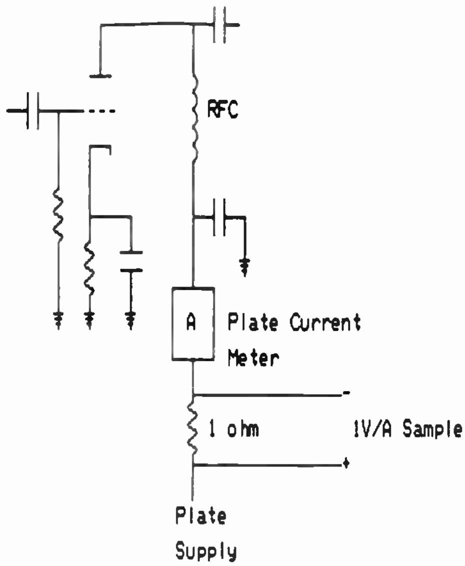


Figure 4.7-4(a). Plate current sample with dangerously high common mode voltage.

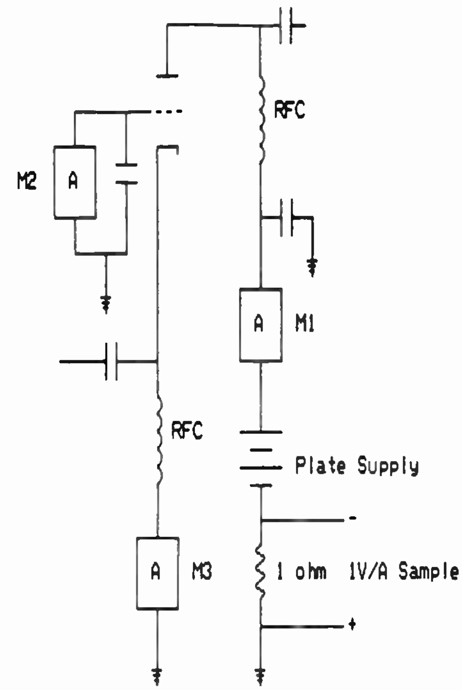


Figure 4.7-4(c). Plate current sampling in negative side of plate supply.

If a high voltage supply is powering only the plate of the RF amplifier (such as the grounded grid amplifier shown in Figure 4.7-4(c)), it is possible to measure the plate current by measuring the current into the negative side of the supply. Instead of grounding the negative side of the supply (as is typically done), a current sense resistor is placed between the supply and ground. The voltage across this resistor is proportional to the plate current. Note that the current through this resistor is the same as the plate current indicated by Ml, while M3 (the cathode current) is the plate current (Ml) plus the grid current (M2). Note also that the sample voltage is negative. Most (though not all) remote control equipment can handle bipolar sample voltages.

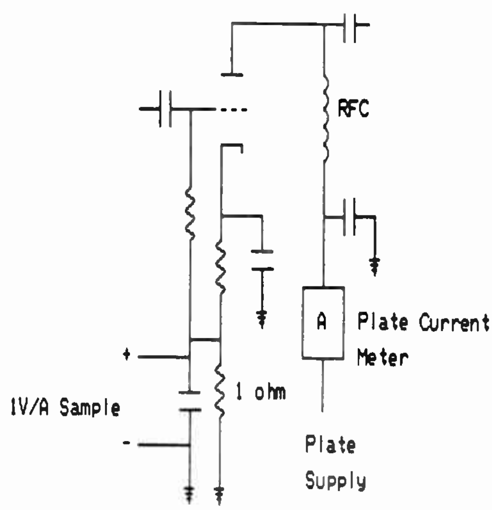


Figure 4.7-4(b). Plate current sampling in cathode circuit. Note that grid is returned above sample resistor.

Some pulse duration modulated AM transmitters operate with the plate of the final amplifier at dc ground potential. The cathode of the final idles at -7.5 kV and peaks at -15 kV on $+100\%$ modulation peaks. The same sampling techniques can be used.

Resistor Stability

The initial tolerance of resistors used as current sense resistors or in voltage dividers is not critical. An inexact resistance will cause the sample voltage to be other than that expected. Whatever sample voltage appears, the remote meter can be calibrated to agree with the local meter. The resistance of resistors changes with time and temperature. The variation with temperature is the major error contributor.

A carbon composition resistor may have a temperature coefficient of about ±600 ppm/°C for resistors up to 1 kΩ. This increases to ±1875 ppm/°C for resistors up to 1 MΩ. If these resistors were used in the plate voltage sampler of Figure 4.7-1(a), the sample voltage *could* change 22% over a 0° to 50°C temperature range. Carbon film resistors are available with temperature coefficients that range from 0 to -300 ppm/°C for low resistance values to 0 to -1000 ppm/°C for high resistance values. Metal film resistors are available with temperature coefficients down to 15 ppm/°C. Bulk metal resistors are available with temperature coefficients of 5 ppm/°C. If the resistors in a voltage divider are at the same temperature, the matching between the temperature coefficient of resistance (TCRs) is more important than the actual TCR. Since the TCR matching is generally better than the TCR of an individual resistor, sampling systems should be designed to keep the resistors in the network at the same temperature.

In differential amplifiers, the matching of resistor values is very critical. To ensure the resistance values track with temperature, it is common to use a resistor network made using hybrid film techniques or integrated circuit techniques. Since the resistors are made at the same time and of the same materials, the TCRs will match closely. Since they are on the same substrate, they will be at about the same temperature.

Other considerations in resistor selection include the power and voltage rating of the resistors. For example, a typical 2 W carbon film resistor is rated at 750 V. A resistor with a higher voltage rating, or several in series, would be needed to sample a high voltage.

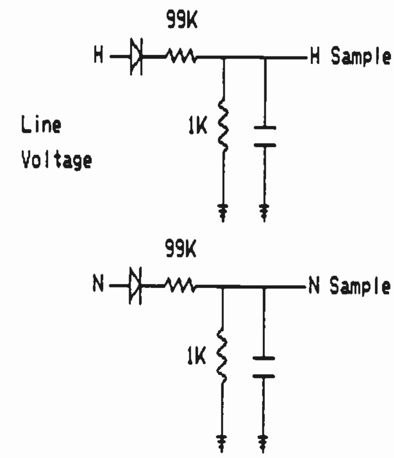


Figure 4.7-5(b). Non isolated line voltage sampler with differential sample output.

AC Voltage Sampling Circuits

Figure 4.7-5(a) shows a simple transformer isolated ac line voltage sensor. The dc output voltage will be approximately proportional to the differential line voltage (as measured between hot and neutral). The linearity of the sample is limited by the diode knee voltage as a proportion of the transformer secondary voltage. This line voltage sampler may be as simple as a calculator supply directly driving the analog input of the remote control.

Figure 4.7-5(b) shows another line voltage sampling circuit. This circuit provides a differential output. Ideally, the neutral line voltage is about zero, so that portion of the circuit may be deleted. If the circuit is used with a remote control that does not have differential inputs, an external differential amplifier may be used, or the hot and neutral samples can drive two input channels of the remote control. The operator or the remote control software can then calculate the actual differential voltage.

Three-Phase Line Voltage Sampling

The *hot* portion of the circuit in Figure 4.7-5(b) can be duplicated for each phase of a three-phase system. The remote control is then presented with four samples, one for each phase voltage referenced to ground, and one for the neutral line voltage. Note, however, that this circuitry will not necessarily indicate a loss of phase. Should a phase open, a load between two phases may pull them together, causing the correct voltage to be on each phase, but the proper phase relationship would not be present. Such an open phase can be detected with the *phase imbalance detector* of Figure 4.7-6.

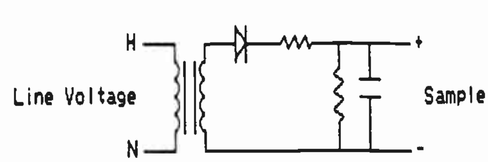


Figure 4.7-5(a). Transformer isolated line voltage sampler.

Figure 4.7-6 shows the circuit driven by a three-phase wye source. The three 10K, 2 W resistors form an averaging circuit, adding the three voltages and dividing the result by three. In a 3 phase wye circuit, Va = 120 V at 0°, Vb = 120 V at 120°, and Vc = 120 V at 240°. These three voltages add up to 0 V at V1, indicating the circuit is balanced.

If a line imbalance occurs (such as a phase being lost where phase B follows phase A), the input voltages are 120 V at 0° on phases A and B, and 120 V at 240° on phase C. Averaging these voltages results in 69.28 V at −30° at V1. The remaining circuitry provides a dc voltage proportional to the half wave average of the voltage at V1. Use of this simple circuit detects loss of a phase while merely measuring phase voltages does not. Note that the standard wiring of a three-phase delta source results in voltages that are not balanced with respect to ground. The typical voltages are 120 V at 0°, 120 V at 180°, and 208 V at 90°. If these are fed to the circuit of Figure 4.7-6, the yield is V1 = 69 V at 90° and a sample output of 13.13 VDC. This

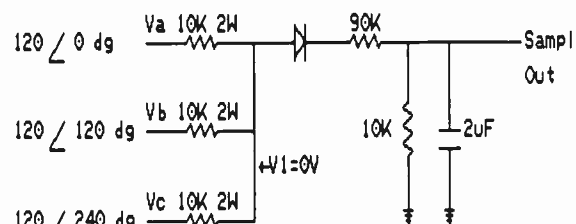


Figure 4.7-6(a). Alternate current line phase imbalance detector. Note that the sample output voltage is 0 volts when ac line is balanced.

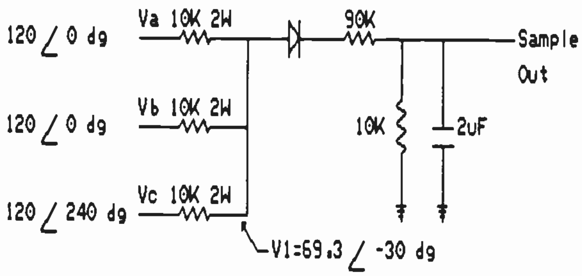


Figure 4.7-6(b). A current line phase imbalance detector with phase B shorted to phase A, giving sample voltage representing imbalance.

sample voltage could be monitored. Significant deviation would indicate a power line problem. Another approach would be to add three transformers between the line and the phase imbalance detector to change the delta source to a wye.

Note that in all of these circuits, the diode is placed before the voltage divider (the 90K and 10K resistors) to minimize the nonlinearity due to the diode knee voltage.

Finally, none of these circuits will detect a phase reversal. The actual phase sequence is generally unimportant to transmitter power supplies. However, many transmitters use three-phase blowers. A phase reversal will cause the blower to rotate the wrong way, severely deteriorating the transmitter cooling. Phase sequence detecting relays that will release a set of contacts on a missing phase or a phase reversal are available. These contacts could drive an analog or a status input of a remote control.

Filament Voltage Sampling

Filament voltage sampling can be done using the same techniques previously described for line voltage sampling. If an isolation transformer is used, the voltage ratio would need to be adjusted to yield the proper sample voltage. If an isolation transformer is not used (instead, just a diode and voltage divider are used), care must be taken to ensure that only the filament voltage is sampled. Grounded grid amplifiers have a high RF voltage on the filaments, so the filament voltage must be measured before the RF isolation inductors. AM transmitters often have a dc common mode voltage on the filament (typically due to circuitry for protective bias, cathode current sensing, and power trim). This can be removed by using an isolation transformer or capacitively coupling the ac voltage to the sampling circuitry.

Tower Light Sampling

The monitoring of tower lighting systems is required by Section 17.47 of the FCC Rules. This section requires a daily observation of either the tower lights or a properly operating tower light indicator by the tower owner. Section 17.47(a)(1) requires the remote indicator (or alarm) to indicate the failure of any lamp on the tower. Section 17.48(a) requires the failure of a

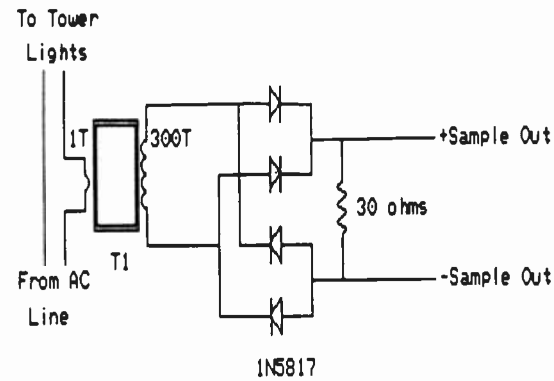


Figure 4.7-7(a). Tower light current sampler provides 84 mV dc average/RMS ac amp.

steady burning lamp at the top of a tower or the failure of any flashing obstruction lamp to be reported to the FAA. A daily observation of tower lights requires that the lights be viewed directly from a location which would allow any lighting failure to be observed. A properly operating tower light indicator is one that will report the proper operation of the lighting system as well as the failure of a required light. Such systems usually operate by sensing the current driving the tower lights. Stations operating by remote control often read this sensed current at the remote control point and interpret proper operation by observing the indicated current.

Figure 4.7-7(a) shows a possible tower light sensor. T1 is a current sensing transformer (Toroid Corporation of Maryland TR-3025) with a primary to secondary current ratio of 300. Ideally, the secondary current is the primary current divided by 300. As circuitry is added to the secondary (secondary not shorted), the secondary current decreases and the linearity of the secondary versus primary current curve decreases. This is due to the output voltage limitations of the transformer (it is not an ideal current source).

Figure 4.7-7(b) shows the peak voltage output of the circuit when measuring the current driving two flashing 620 W lamps and two steady 116 W lamps (all rated at and driven with 130 VAC). Note the high

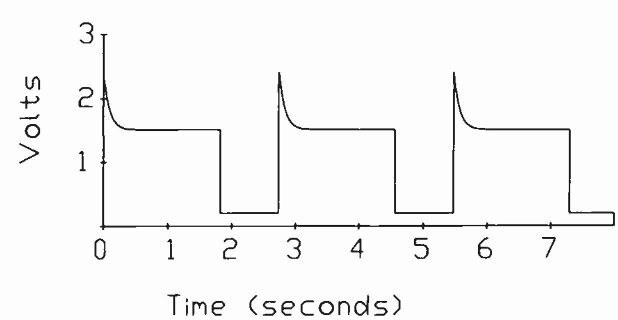


Figure 4.7-7(b). Tower light sample output waveform.

output when the beacon is first lit. This high inrush current (due to the filament cold resistance) increases the initial beacon current by about 70% (generally limited by the resistance of the wiring feeding the lamp). The time constant for the exponential decay down to the steady state current is about 91 ms. If the remote control system uses an integrating (dual slope) analog to digital converter, it will measure the average dc voltage across the resistor during the sample time. A typical dual slope A/D does between 3 and 30 conversions per second. It is not sampling the input during the total conversion time (time is spent in the autozero and reference integrate phases). Assuming 3 conversions per second, 2.73 conversions can be done during the beacon off time (using the typical flash rate and duty cycle) and 5.46 conversions can be completed during the beacon on time. By watching the minimum and maximum conversions, it is possible to determine if all lamps are working properly. Many people will be tempted to add a capacitor across the output of the sampler to provide a dc sample to the remote control. This, however, will cause samples taken during the beacon off time to include a component due to the beacon on current and vice versa. This interference between the measured on and off currents is eliminated by allowing the A/D integrator to remove the ac line frequency (actually twice line frequency, since it is full wave rectified) component, though there is still some interaction possible due to a conversion starting when the lamp is on and completing when the lamp is off. A/D sampling rates on dual slope converters are typically chosen to minimize the effects of 60 Hz and its harmonics.

Table 4.7-1 shows typical minimum, maximum, and average for a series of samples taken over a one minute period with 3 conversions per second. Note that these voltages are less than the peak voltages shown in Figure 4.7-7(b) (measured with an oscilloscope), since the voltages in Table 4.7-1 are average dc voltages, measured with an integrating A/D converter.

Note that the circuit of Figure 4.7-7(a) measures tower light current, which will vary with tower light voltage. FCC Rule Section 17.54 requires the voltage at the base of tower lamps during normal operation to be at least 97% of the rated lamp voltage. Tower lamps tend to be constant current devices. As the line voltage drops, the filament cools, decreasing the filament resistance and increasing the current, though not back to the original value. With a typical tower lighting load (two 620 W lamps and two 116 W lamps) a 3% decrease in line voltage causes a 1.57% decrease in lamp current. A 6% decrease in line voltage causes a 3.30% decrease in line current. Since these variations due to line voltage fluctuations (staying within FCC required line voltage limits) are much less than the variations due to a lamp failure, it is simple to differentiate between lamp failures and permissible line voltage variations.

Table 4.7-1

Average dc voltages (as opposed to RMS) from circuit of Figure 4.7-7(a) under various tower lamp failures, 3 samples per second for 60 seconds.

Light Condition	Minimum Volts	Maximum Volts	Average Volts
All ok	0.132 (100%)	1.402 (100%)	0.766 (100%)
Side light out	0.060 (45.5%)	1.217 (86.8%)	0.687 (89.7%)
Beacon out	0.132 (100%)	0.776 (54.6%)	0.465 (60.7%)

If a remote control returns voltage indications corresponding to tower light current (using a circuit similar to that suggested), minimums and maximums can be checked by an operator or system software to determine if the tower lights are operating properly. If the output of the circuit is run through a low pass filter with a cutoff frequency well below the beacon flash rate, a steady dc voltage corresponding to the average voltage would be available. This voltage is proportional to the current drawn by the steady burning lamps plus the product of the current drawn by the flashing beacons and the duty cycle of the flashing beacons. A variation of more than 5% in this voltage would indicate a lamp failure.

All the discussion has been based on a standard tower lighting load. This consists of two flashing 620 W lamps and two steady 116 W lamps. Tall towers require more lamps. It is suggested that the lighting circuits for such towers be sampled individually, breaking the system down to several standard lighting loads. Increasing the load substantially will make it difficult to detect the loss of one small lamp.

No matter what approach is used to do tower light sampling, it is suggested that a single 116 W lamp be added to each lighting circuit. This light should remain on at all times or switch on when the other tower lights are turned on. For AM stations, this lamp might be in the antenna tuning unit at the base of each tower. FM and TV stations might use it as a transmitter site nightlight. The tower light sensing circuitry and software can be tested by turning off this lamp, simulating the failure of a steady burning lamp. An operator, system hardware, or system software should be able to detect this tower lamp failure. This is an easier method of testing the system than climbing the tower to unscrew a lightbulb!

The FCC permits the use of high intensity (strobe) lighting as an alternative to tower painting and standard lighting. Due to the complexity of the required control circuitry, strobe control systems typically provide outputs that can be used to determine the state of the system (high/medium/low intensity and failure of any lamp). These outputs can be returned to the transmitter operator using standard status monitoring circuitry.

In addition, several manufacturers offer tower light alarm systems for incandescent tower light systems.

FM AND TV OUTPUT POWER SAMPLING

FM and TV transmitters are required to have a metering circuit that measures the output power. Most transmitters have a remote output of this sensor that can be used to drive remote meters. These meters are generally directional couplers that sample the forward and reflected power at the transmitter output. For those transmitters without a remote output power sample (or those stations using combined transmitters), a transmission line directional coupler with an associated wattmeter (with remote sample) can be added. NTSC television transmitter output power meters include a peak detector circuit, since the RMS output power varies substantially with picture content. The NTSC peak amplitude (tip of sync) should be a constant power. Different power measurement techniques are used with the newly adopted digital television standard.

The dc output of the reflectometer has a relatively high source impedance and is easily loaded. Many transmitters include isolation circuitry so the remote or extension meter does not load the directional coupler, causing inaccurate indications on the transmitter front panel meter.

The sample voltage out of a directional coupler is proportional to the square root of the power (directly proportional to the voltage or current), except at very low powers, where the diode knee voltage again causes additional nonlinearities. The output power meter on most transmitters is a mechanical meter with a nonlinear scale, allowing it to read output power directly. An analog remote meter can also use the same scale, allowing direct reading of the transmitter output power. Digital meters, however, generally use a linear A/D converter. Getting an accurate remote indication involves squaring the sample voltage either before or after the A/D converter. The sample voltage can be squared prior to the A/D using an analog multiplier (or balanced modulator) and tying the two inputs together (multiplying the sample voltage by itself). Devices that handle this analog squaring are often called *power to linear converters*.

The sample voltage can also be squared after the A/D converter. This may be done by giving the transmitter operator a chart (indicated power versus actual power). Finally, most microprocessor based remote controls make this calculation automatically.

AM POWER SAMPLING

AM stations generally determine power by measuring the RF current into the antenna system. For nondirectional stations, this is the base current of a series fed antenna or the feedwire current of a shunt fed antenna. A directional station measures the current into the common point of the array. In each case, the measured current is squared and multiplied by the resistance (base, feed wire or common point) to arrive at the power into the antenna. Section 73.51(a)(1) of the FCC Rules also permits the use of direct reading power meters that measure the voltage, current and phase relationship between them to determine the power, but these meters are rare.

RF currents have traditionally been measured with thermocouple meters, which measure the true RMS current. Remote thermocouple meters are available, but suffer from nonlinearity between the RF current and the sample voltage. Further, when an AM transmitter is modulated, the RMS current into the antenna increases (corresponding with the addition of power to the AM sidebands). The FCC places limits on the unmodulated antenna current. Having the antenna current indication vary substantially with modulation would require checking the antenna current without modulation. For these reasons, diode meters are now generally used for remote RF current indications.

Diode meters utilize an RF current sensing transformer to develop a dc voltage that is proportional to the RF current. The dc sample is obtained by running the rectified RF through a low pass filter that removes the ac components (RF and audio). Ideally, the dc component of the rectified RF is constant, since the transmitter ac couples the audio into the final amplifier (using typical modulation techniques). The sample voltage will be the same with or without modulation, allowing measurement of the unmodulated antenna current without interrupting modulation. Again, there are potential nonlinearities due to diode knee voltages. However, these may be overcome by placing the current sense transformer terminating resistor after the rectifier, as was done in the tower light sensor previously described. Other techniques are available to eliminate this nonlinearity. These include developing a bias voltage to get over the knee and the use of zero crossing driven FET switches (similar to a synchronous rectifier).

A well designed diode meter (such as the Delta TCA series) also serves well as the local antenna current meter. These meters provide a local indication and a sample voltage suitable for driving remote metering.

In practice, the indication of a diode type RF ammeter will vary with modulation. This is generally due to less than perfect carrier amplitude regulation (carrier shift) in the transmitter. Often the transmitter high voltage power supply is loaded by the modulators as the modulation level is increased, decreasing the high voltage available for the final amplifier. This can cause a decrease in the indicated antenna current. Variations in the antenna impedance with frequency may also cause carrier shift, as the various sideband frequencies see a varying load impedance. If these variations in indicated antenna current are excessive (remote antenna current meters are required to agree with the local meter within 2% [Section 73.57(d)]), the readings would have to be taken without modulation, or perhaps some electrical compensation could be added (see *Carrier Shift Compensation*).

Some stations take a sample of the RF voltage at the antenna input (base, feed wire or common point) and calibrate the remote indication to agree with the local ammeter. Assuming the impedance of the antenna is constant, then the RF voltage will indeed be proportional to the current. This approach uses the antenna itself as a *current to voltage converter*.

Finally, directional stations are equipped with a very high quality RF detector circuit in their antenna monitor. Stations that operate nondirectional some portion of the time may use the indication of the antenna monitor current sample for the nondirectional tower as a remote base current sample. The actual RF sample is typically from an RF current transformer at the base of the tower or a current sampling loop part way up the tower. Stations can also add a current sense transformer just prior to the common point meter. This transformer can drive the antenna monitor to give a remote sample of the common point current.

Carrier Shift Compensation

The circuit of Figure 4.7-8 may be added between a Delta TCA RF ammeter and a remote control (input resistance of 1 M or MΩ greater) to compensate for variations in indicated antenna current due to carrier amplitude shift with modulation. Figure 4.7-9 shows the remote sample voltage out of a Delta TCA2OEXR with modulation. Normally, the ac component (due to modulation) is removed by a low pass filter in the remote control or the *mechanical low pass filter* created by the movement of mass in a meter movement. If R2 is set to the center of its range, the circuit of Figure 4.7-8 forms a low pass filter (Fh = .034 Hz). The output voltage will be the dc component of the input, regardless of any ac component (due to modulation). If, during modulation, the dc component drops (due to carrier shift), the dc output voltage will drop. If, however, the wiper of R2 is moved towards Dl, Cl will tend to charge to the peak input voltage instead of the average voltage. If the output voltage without modulation is measured (remote control calibrated without modulation), then R2 adjusted for the same indication with modulation, most of the effects of carrier shift can be removed. Testing of the circuit on a 5 kW plate modulated transmitter operating into a three tower directional array reduced variations in indicated common point current from ±1.2% to ±0.1%.

Note that the circuit must not be loaded by the remote metering circuitry. If a low resistance (less than 1M or MΩ) is to be driven, a standard operational amplifier voltage follower circuit should be added between the carrier shift compensator and the remote metering system.

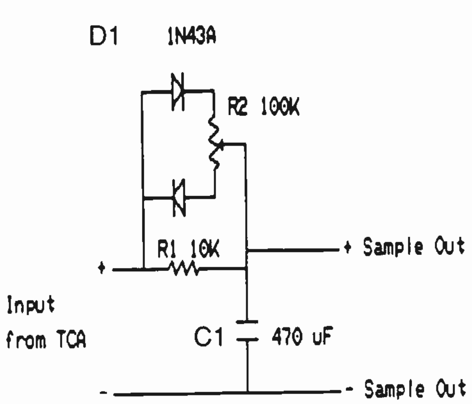


Figure 4.7-8. Remote ammeter carrier shift compensator.

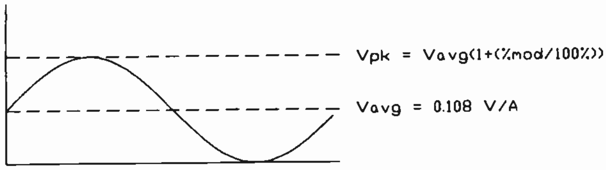


Figure 4.7-9. Remote sampler output waveform of Delta TCA20EXR. Note that Varg will typically decrease with modulation due to carrier shift. Vpk increases with modulation.

DIRECTIONAL ANTENNA MONITOR

AM radio stations with directional antenna systems must have an antenna monitor to verify the proper operation of the antenna system when operated in a directional mode. The indications of a directional antenna monitor may also drive remote meters when designed for such use and when authorized by the FCC (Section 73.53(a)(9)). Most antenna monitors require a tower select input and usually a pattern select input in order to measure the amplitude (or ratio) and phase of the selected tower, though some more recent monitors provide continuous outputs of all samples. Additional difficulties are caused by the necessity to properly indicate phase relationships that may be positive, negative or very close to 0°. In extension metering systems, the tower select controls can be extended to the metering point, allowing the operator to select which tower the monitor will measure.

Most (though not all) antenna monitors use one or two amplitude detectors and one phase detector. This technique requires that the detector circuitry be shared among several towers through the use of an RF multiplexer which selects the desired tower input samples. Monitors with two amplitude detectors leave one connected to the reference tower at all times. The other (or only) amplitude detector is switched between the RF samples from the various towers in the antenna system. One input of the phase detector is also constantly connected to the reference tower. The other input of the phase detector follows the amplitude detector, connecting to one tower sample at a time. When a tower is selected on such an antenna monitor, the response time of the RF multiplexer and the settling time of the detection circuitry prevent an accurate indication from being immediately available. At least one antenna monitor avoids this problem by providing complete detector circuitry for each tower input sample. This approach eliminates the need for tower input selection and can provide continuous indications of all ratios and phases to remote control equipment.

In remote control applications, when tower select inputs are required, the control outputs of the remote control system are typically used to drive the tower select inputs of the antenna monitor. A fairly common practice has been to assign a tower to a channel, then use the *raise* control to read the phase and the *lower* control to read the loop current or ratio. This technique, however, requires interface circuitry between the remote control and the antenna monitor (the control contacts are selecting which indication, phase or ratio, as well as the tower to be sampled) and may result in calibration difficulties, since two dissimilar readings will need to be calibrated on a single analog telemetry input. A simpler approach is to have one or more channels assigned to phase and one or more channels assigned to loop current. The analog inputs of these channels can connect directly to the antenna monitor phase and loop current sample outputs, and each channel will be calibrated to only one type of indication. The *raise* and *lower* control lines then drive the tower select inputs of the antenna monitor. Table 4.7-2 lists some sample channel assignments for such a system.

This arrangement removes the requirement for interface relays between the antenna monitor and the remote control. The current indications may be sample currents (typically from sampling transformers at the tower base or a sampling loop on the tower) or may be current ratios.

Some remote controls provide a channel selected output that notifies external sampling equipment that the channel is selected. With such a system, the channel selected output can directly drive the tower select input of the antenna monitor, allowing the operator to get an antenna monitor reading in the same manner as any other reading (by simply selecting a channel). Such systems can also scan the monitor channels automatically. In this case the response time of the RF multiplexer and settling time of the detector circuits must be considered. Most antenna monitors will present a valid sample about one second after a tower is selected.

To simplify the remote interface of antenna monitors, *sample and hold* devices are available. These devices scan through the tower selects on the antenna monitor and store the resulting sample voltages. The device then provides a steady output sample voltage for the phase and loop current (or ratio) for each tower in the system. This makes the antenna monitor appear like other samples at a transmitter site (continuously available). Operators and system software must allow for the possibility that such a sample and hold system may have acquired some data with the station operating in one pattern, and the remainder of the data in another pattern. This is also a possibility in any system that scans the antenna monitor parameters. Some system software checks the pattern select (day/night) before scanning the antenna monitor and again after scanning the antenna monitor. If these samples of the pattern select disagree, the monitor is scanned again.

Table 4.7-2
Suggested channel assignments for remote monitoring of a 4-tower DA.

Parameter	Control (channel)
Tower 1 phase	Raise (10)
Tower 2 phase	Lower (10)
Tower 3 phase	Raise (11)
Tower 4 phase	Lower (11)
Tower 1 current	Raise (12)
Tower 2 current	Lower (12)
Tower 3 current	Raise (13)
Tower 4 current	Lower (13)

Getting a remote indication of the sign of the phase angle indication on some antenna monitors is difficult. Most antenna monitors output a positive voltage (typically 0 to 1.800 V for 0° to 180°), whether the sign of the phase is positive or negative. Some antenna monitors introduce a delay in the reference tower circuitry when a phase sign switch is pressed. This delay will cause an increase in the phase indication if the phase is positive (selected tower leads reference tower), or a decrease in the phase indication if the phase is negative. Although this switch could potentially be remote, it rarely is. Some antenna monitors use digital techniques (such as a D flip-flop) to detect the sign of the phase. While the phase sign is brought out for remote display (typically as a TTL level), the analog output for driving the remote control still outputs a positive voltage regardless of the phase sign. This phase sign logic level can be returned to the operator through a remote control status channel. The operator (or system software) can negate the phase indication, when appropriate.

Since it is rare to have the sign of a DA tower phase change (unless it is very near 0° or 180°), many stations wire the antenna monitor to the remote control such that the proper sign is indicated (reverse input leads to get a negative reading), if the remote control has a floating analog input. Others may merely note the sign of the phase for each tower in the operator instructions at the remote control point. Microprocessor based remote controls may use a negative calibration or scaling factor to cause a positive input voltage to display a negative value (corresponding to a negative phase).

Finally, note that most antenna monitors will require a *pattern select* control input. This input may change reference levels and/or reference towers. This input is typically driven by the antenna phasing equipment control system. This insures that the antenna monitor and the antenna system agree on the pattern.

PROBLEMS IN ANALOG SAMPLING INTERFACING

Although standards for interfacing transmitters and monitoring equipment exist (see IEC 60864), they are not widely used in the U.S. One cannot always connect the remote outputs of a transmitter directly to a remote control. A few typical problems and suggested solutions are listed here.

Sample Voltage Too High

To eliminate errors contributed by voltage dividers inside a remote control, many manufacturers run the sample voltage directly through an analog multiplexer into the A/D converter. Voltage limitations on the multiplexer and the A/D converter limit the maximum sample voltage the system will accept. Exceeding the multiplexer maximum voltage will often introduce leakage into other channels, making all readings appear to be out of calibration. Exceeding the A/D maximum input voltage will cause an erroneous reading on that channel (or an overrange indication) and may cause erroneous readings on other channels (due to A/D overload recovery time).

Some transmitters provide adjustable sample voltages. In this case, the sample voltage should be adjusted down to match the optimum input of the remote control.

Sometimes a voltage divider can be modified by adding an external resistor across the remote output of the transmitter, reducing the sample voltage. This will not be the case if the transmitter includes an isolation amplifier for the metering circuitry. The added shunt resistor should be of high quality with a low temperature coefficient of resistance (TCR).

The sample voltage output of the transmitter can be run through another voltage divider, reducing the sample down to the optimum for the remote control. Again, high quality, low TCR resistors should be used.

Sample Voltage Too Low

Too low a sample voltage will cause a lack of resolution in remote metering circuitry. Many systems will not allow calibration if the sample voltage is less than 100 times the A/D step size, since this would give a resolution of less than 1%. The only solution to low sample voltages is the addition of an amplifier between the sample and the remote control. This may be required on FM or TV power meters (reflectometers or directional couplers).

Some parameters are ideally zero (such as reference tower phase, three-phase line imbalance, or reflected power). The sample voltage for these parameters are also ideally zero, making it difficult to calibrate remote metering. If the antenna monitor has a *180° calibrate* button, this switch can be held while the remote control is being calibrated. The system is calibrated to the indicated 180°. When the calibrate button is released, the remote control should agree with the local meter.

A three-phase balance detector can be calibrated by purposely unbalancing the three-phase input. This is typically accomplished by disconnecting one input of the detector and grounding it (setting it at 0 V). The unbalance can be measured with a voltmeter. The remote metering can then be calibrated to agree with this indication.

The remote indication of a reflected power sample can be calibrated by calibrating it to the forward power. If the transmitter uses the same circuitry (same size power sampler) for forward and reflected power, the input of the remote control can be temporarily connected to the forward power sample. The remote metering is then calibrated to agree with the forward power indication. The remote control input is then reconnected to the proper point.

Some transmitters utilize rotatable *slugs* to sample the forward and reflected power. The reflected power slug can be rotated to measure forward power. If the reflected power slug is more sensitive than the forward power slug (full-scale reflected is less than full-scale forward), then the transmitter power should first be trimmed down to the maximum the reflected sampler can handle. If the transmitter includes VSWR shutdown circuitry, it will have to be disabled during this adjustment. Otherwise, the transmitter will immediately shut down.

Similarly, some transmitters have coaxial cables going out to external forward and reflected power samplers. These cables can be temporarily reversed, giving a high reflected power indication, allowing calibration. Again, VSWR shutdown circuitry would have to be disabled during this time.

Sample Voltage Differential, Remote Single Ended

Often a transmitter provides differential sample voltages while the remote control expects one side of a sample voltage to be grounded. There are various techniques for handling this. One solution (the historically typical approach) is to add a differential amplifier between the transmitter and the remote control. Another solution involves measuring each side of the differential voltage with respect to ground, then determining the differential voltage through calculation (either an operator calculation or a computer calculation).

Sample Voltage Not Linear

Generally, the sample voltage is the measured parameter multiplied by a constant. Sometimes it is the square root of the measured parameter (typically for power indicators). Other times it may be linearly related to the parameter, but offset (zero sample voltage does not correspond to zero parameter, often the case for temperature indicators).

In each case, the calculation required to convert the sample voltage back to the parameter can be done before the remote control, in the remote control or after the remote control. *Power-to-linear converters* can square the sample voltage so it linearly tracks the power. Differential amplifiers often include an offset control that can be used to subtract whatever may be necessary to make the sample voltage track the parameter properly. The remote control itself may contain calculation software to perform these calculations after the A/D conversion. Finally, the remote control may just indicate the sample voltage. The operator may use a calculator or a look-up table to determine the actual parameter.

STATUS SAMPLING

It is often desirable to remotely indicate the status of on/off indicators. Many transmitters utilize back lit push button switches for local control and status indication. The switch that corresponds to the current status is lit. For example, if there are a pair of switches marked *plate off* and *plate on*, pressing the *plate on* switch will cause the indicator in the *plate off* switch to extinguish, while the indicator in the *plate on* switch will be illuminated. This sort of indication is very user-friendly. An operator could determine that the plates had been turned on by looking at the plate voltage or current meter, but seeing that the button lit is a more intuitive indication that the desired action took place.

Other status indications that are often available at a transmitter site include filaments on/off, power high/low, stereo/mono, overload tally, RF switch settings (including DA pattern select relays, transmitter select relays, and combiner relays), STL carrier presence, STL selected, audio processor selected, fire alarm and burglar alarm. Each of these is an on/off indication. Many are isolated switch contact closures. Some are switch (or open collector transistor) closures to ground. Some are switch (or open collector) closures to some supply voltage. Most remote controls accept closures to ground. If the status input is open, the remote control pulls the status line to +5 V. This input design is generally used due to low cost. The inputs drive complementary metal oxide semiconductor (CMOS) or transistor transistor logic (TTL) inputs through input protection circuitry. If a transmitter status output is not compatible with the remote control status inputs, additional interface circuitry may be added. This circuitry may consist of relays or optical couplers, which provide complete isolation between the transmitter and the remote control. Resistor/transistor circuits or integrated circuits may also be used for *level translation*, converting the transmitter's status levels to levels appropriate for the remote control.

Finally, the transmitter status outputs can drive analog inputs to the remote control. These analog samples can be compared with a threshold (by the operator or system limit checking software) to determine whether the status is true or false. Use of analog inputs with appropriate software support allows *universal inputs* that can serve as either analog or status inputs. Note that a status input usually includes a *pull-up* resistor (so the input can detect a closure to ground), while analog inputs do not. A universal input system can provide a single pull-up after the analog multiplexer and enable or disable the pull-up under program control.

CONTROL OUTPUTS

Transmitter remote controls have traditionally provided momentary isolated relay contact closures for control. Two control outputs were provided for each analog input. These outputs were designated *raise* and *lower*. If the operator had analog channel 1 selected (perhaps plate voltage), pressing the *raise* button closed a pair of contacts that were wired to the plates on control input of the transmitter. The transmitter plates would go on, and the operator would see the plate voltage go up. Pressing *lower* while plate voltage was selected turned the plates off. If channel 2 were selected (perhaps plate current), *raise* and *lower* might change the transmitter between high and low power.

Use of momentary controls allows the remote control to be connected in parallel with the local control (transmitter front panel) while allowing control from both locations simultaneously. Latched control lines (non-momentary) would disallow control from more than one location.

Use of isolated relay contacts allows the remote control to drive a variety of control circuits. Older transmitters used 120 to 240 V ac to drive the control relays. Isolated relay contacts were suitable for driving these transmitters, though such transmitters did place hazardous voltages on the remote control. Some current transmitters are expecting the remote control to provide a momentary contact closure to some positive voltage (generally +12 or +28 V). Again, remote controls with isolated relay contacts can drive these directly. Many remote control systems provide open collector control outputs. This is generally a cost saving technique. A single integrated circuit may provide four, eight or even 32 open collector outputs. This compares quite favorably to one control output per relay. Further, open collector outputs that all switch to a common ground also allow a higher wiring density. Figure 4.7-10 shows the interconnection of a remote control open collector output to a transmitter control input. The diode across the relay coil clamps the control voltage when the transistor turns off, preventing the transistor from breaking down due to excessive voltage. A remote control may include clamp circuitry to protect the control output transistors.

If a transmitter uses high side control switching (expecting a closure to a positive voltage), it may be possible to reverse the power supply leads to the control interface portion of the transmitter, converting it to ground switching. This eliminates the need for repeat relays between a remote control with open collector outputs and the transmitter. If the control relays include clamp diodes across the coils, they will have to be reversed. Further, if the transmitter uses magnetic

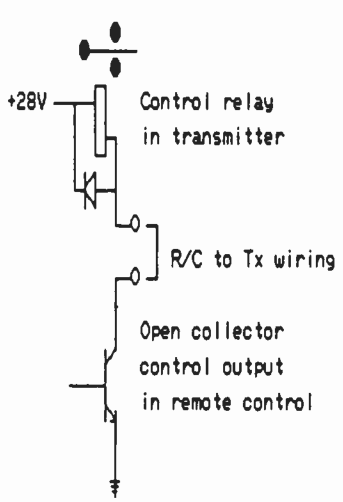


Figure 4.7-10. Open collector remote control output driving a transmitter control relay.

latching relays in the control circuit, their actions will probably reverse due to the change in current direction. Finally, at least one transmitter expects contact closures to ground, but interrupts the ground with the local/remote switch. If the remote control uses relay outputs to switch the control inputs to the switched ground of the transmitter, all works fine. However, if the remote control switches the control lines directly to ground (as an open collector output would), the remote/local switch is bypassed. This presents a safety problem to someone unfamiliar with the transmitter site. The remote control would still be able to turn on the transmitter, even though the local/remote switch was in local. It may be possible to move the wiring of the remote/local switch to the top of the control relay coils. The switch would then interrupt the positive supply to the relays when the switch was in local.

It would seem, from a remote control manufacturer's point of view, that all control and status should rely upon closures to ground via mechanical contacts or open collector (or drain) transistors. However, transmitter manufacturers point to safety advantages in using high side switching in control circuits. It is, as they point out, far less likely for a control wire to accidentally short to +24 V than it is for that wire to short to ground, possibly activating a transmitter. Further the standard IEC 60864 requires control outputs to be "earth free." (See the section on IEC 60864.)

Control Interface Panel

In all but the simplest transmitter sites, it is common to have a relay interface panel between the remote control and the various pieces of transmitter site equipment. Besides any repeat relays that may be required to convert from open collector drive to whatever the transmitters may require, such a relay interface panel will generally use relay ladder logic to interlock various transmitter site functions. Such an interface panel may handle the sequencing of RF switches in transmitter or DA switching, and ensure that a transmitter drives a load before it can be enabled and that dummy load cooling is on before power is sent to a dummy load.

In the design of a transmitter site, there is an interesting division of responsibility problem. Various portions of the site control and monitoring can be assigned to the remote control, control interface or the transmitter. It has been common practice to have a transmitter accept control pulses and report its current status back to external equipment. Many transmitters now include an automatic power control, which moves some of the control into the transmitter. Some other transmitters have moved much or all of what has typically been considered a remote control's job into the transmitter itself. In most stations, however, there is a fair amount of decision making required outside the transmitter. A fairly common decision problem is the RF sequencing of an AM station pattern change. Similar sequences are required for transmitter changes in other AM stations, FM stations and TV stations.

AM Pattern Change Sequencing

Many stations have used time delay sequencing circuitry to do a pattern change while insuring the RF relays did a *cold switch* (switching while no RF present). A typical sequence might be:

1. Drop transmitter carrier.
2. Wait one second.
3. Switch RF relays.
4. Wait one second.
5. Bring up transmitter carrier.

There is another approach that results in faster pattern changes and ensures the station will not come back up if all the RF relays do not operate properly. This interlocking of DA RF relays was an FCC requirement (under the operator requirement rules) when the FCC first reduced operator requirements for DA stations to less than a First Class Radiotelephone Operator. The interlocking rule has been deleted, but is still good practice (ensuring against operation with other than the desired pattern, and protecting equipment from excessive currents or voltages should a relay stick). An interlocking and sequencing system can be easily setup using relay ladder logic in a remote control interface panel. In relay ladder logic, logical *and* gates are formed by wiring relay contacts in series, while logical *or* gates are formed by wiring relay contacts in parallel. It is also fairly common to form *or* gates using diodes in combination with the relays, reducing the number of contacts a relay may require. The logic for a typical DA pattern change is listed below.

```
IF NightSelected AND OnDayPattern AND PlatesOn
THEN CommandPlatesOff
IF NightSelected AND OnDayPattern AND PlatesOff
THEN RFSwitchToNight
IF NightSelected AND OnNightPattern AND PlatesOff
THEN CommandPlatesOn
```

This represents three series strings of relay contacts.

NightSelected is the pattern change command from the operator. OnDayPattern and OnNightPattern represent the current condition of all the RF relays. Neither will be true during the switch (as the relays change the system configuration). The PlatesOn and PlatesOff status lines correspond to the transmitter's state. Actually, operators are interested in whether there is any RF in the system. An interface panel may use an RF sample from the final amplifier of the transmitter (often a modulation monitor tap) to drive a relay that indicates whether RF is present or not.

Some AM transmitters have a very fast carrier interrupt input. This on/off input drives the modulator to 100% (envelope voltage zero) in pulse duration, digital and series modulated transmitters. On receiving a carrier drop command, the transmitter drops the carrier in 100 μsec or less. Since it takes substantially longer than this for an RF relay to open a contact, the carrier drop and relay drive signals can be supplied simultaneously.

The carrier interrupt input on some pulse duration modulated transmitters expects a floating short (in a switched +28 V dc line) to keep the carrier up, and an open to drop the carrier. This can be driven with floating relay contacts. In many cases, a resistor (10 kΩ typical) may be added across the carrier interrupt terminals. The carrier drop circuit input (the side of the resistor that is driven by the +28 V, not the side that provides the voltage) may be grounded with a relay contact or an open collector driver to interrupt the carrier.

Finally, in the design of the control system for AM directional stations, it is suggested that the relay interface panel send low voltage dc (typically 24 V) to repeat relays in the tuning units at the base of each tower. Standard 25 pair telephone cable can be looped past each tower. This cable carries these low voltage control signals. It also carries the status of each RF contactor in each antenna coupling unit. The cable can also be used to send sample voltages for RF base currents, tower light sample currents and to provide a telephone at the base of each tower. The pattern change repeat relays switch the coil voltage to the RF contactors. The contactors can be powered by the same wiring used to drive any tower lights, reducing the requirements for heavy wiring. It is suggested that the contactors be run on the highest voltage available (typically 208 or 240 V), reducing the current requirements and the resulting voltage drop in the power wiring.

IEC 60864

This standard (available from the International Electrotechnical Commission at http://www.iec.ch) covers the *Standardization of Interconnection between Broadcasting Transmitters or Transmitter Systems and Supervisory Equipment.* The standard is broken into two parts—part 1 covers *dedicated interconnections* (similar to that discussed above), and part 2 discusses a data bus for interconnecting equipment.

With this standard, a transmitter supervisor box is driving a system logic unit that actually drives the transmitters and changeover relays. The *transmitter supervisory equipment* would be a remote control, while the *system logic unit* could be a relay interface panel. Part 2 of the standard puts a bit more intelligence in the system logic unit than can be accomplished with relays.

Even though part 1 of the standard was adopted in 1986, it is not widely used in the U.S. The standard allows the use of relays, semiconductors or optocouplers for the control or status (binary) interface. When a relay interface is used, commands may be presented as momentary (100 to 500 ms) closures or continuous closures (latched). Indications (status) lines are to remain activated as long as the status is true. The standard states that the circuits shall be *earth free* (ungrounded) and may be independent pairs or utilize a common return wire (which, again, must not be grounded). These circuits utilize a nominal open circuit voltage of 24 V dc and have a maximum operating current of 25 mA.

When semiconductor or optocoupler techniques are used, the output device (whether a control system command output or a transmitter status output) must be *earth free*, handle a nominal 24 VDC open circuit voltage and a nominal 8 mA closed circuit current (with a maximum voltage drop of 3 V). The circuitry must withstand the transient of a 33 μF capacitor charged to 250 V being discharged into the circuit.

For analog samples (though analog signals *may* be used for control as well as telemetry), sample voltages are specified as being 0 V to +10 V. The minimum load resistance is to be 1 kΩ. The maximum source resistance is to be 100 Ω. The maximum positive or negative common mode voltage is to be 25 V. The standard also requires analog interface terminals to be *earth free*, requiring the use of isolation amplifiers. Analog wiring between supervisory equipment and the transmitters is to be done with shielded pair cables.

The standard also allows for a current loop analog interface. This interface is standard in industrial control, but is rarely used in the U.S. broadcast industry. With this standard, an indication of zero is represented by a current of 4 mA. Full-scale is represented by a current of 20 mA. A current sensor is allowed to have a maximum resistance of 1 kΩ. The current generator is to have a high resistance making the loop current independent of load resistances.

Part 2 of the standard (adopted in 1997) establishes a digital interface standard between supervisory equipment and a *system logic unit.* The system logic unit provides the actual hardware interface to multiple transmitters at the site using discrete wire techniques (as specified in part 1 of the standard). The digital interface is based on the IEEE 1118 *bit bus* standard with the IEC 8482 electrical specifications. The standard provides for a wide variety of digital codes for commands, telemetry and status messages.

As mentioned previously, this standard is not widely used in the U.S. Earth free requirements increase system costs (due to the additional cost of relays and isolation amplifiers) and decreases system wiring density. As the industry becomes more international, this equipment will probably be introduced to the U.S.

REMOTE CONTROL HARDWARE DESIGN

The basic purpose of a transmitter remote control system is to extend the transmitter metering and control from the transmitter site to the operator. This may be accomplished by running a pair of wires for each meter and another for each control. This is the technique used in extension metering. It can be used in remote control also, but it is generally more economical to combine several functions onto a single pair of wires. The basic transmitter site remote control unit multiplexes the metering, status, and control signals onto a single communications circuit. This circuit may be a dc circuit, voice grade circuit or digital circuit.

A dc circuit is a pair of wires that has dc continuity from one end to the other (a *metallic pair*). Direct current and low frequency ac (typically to 3 kHz) may be sent through these circuits. Early broadcast remote controls used two dc circuits one for control and the other for metering. The metering sample selected by the operator is put directly on the metering pair. A current or voltage meter at the control point displayed the transmitter parameter. The various control functions could be sent over a single control pair by feeding various levels and polarities of voltages between the control conductors, or between each conductor and ground.

Later single pair remote control systems utilized audible tones for control and continued to use dc for metering. This allowed both signals to be on the same pair of wires using a form of frequency division multiplexing.

Voice grade circuits typically pass a 300 Hz to 3 kHz band of frequencies. Voice grade circuits are available over a very wide range of media including telephone lines, radio links, subcarriers on microwave, broadcast and satellite systems. Most broadcast remote control equipment is designed to work over voice grade circuits, since they are so widely available. Many signaling techniques are available using voice grade circuits.

Digital circuits are circuits that present a digital interface to the customer provided equipment, at each end of the circuit. Between the customer terminations, the circuit may have taken any of several forms. It may have been run through modems, converting the data for transmission over a voice grade circuit. It may have been combined with other digital data into a high speed data stream for transmission over microwave, satellite, or fiber optic cable. At the other end of the high speed medium, the data are split off to each customer. A remote control must combine the various control and telemetry signals, then encode them in a manner suitable for transmission over the medium available.

Analog Multiplexing

Since several sample voltages are generally sent over one available circuit, a switch is used to select the sample voltages, one at a time, for transmission to the control point. In many remote control systems, this switch is under the direct control of the operator. When the operator selects a metering channel, the switch at the transmitter site selects the appropriate sample and sends the result to the control point for display. These systems may be thought of as a voltmeter with very long leads. Just as a technician moves voltmeter probes from point to point in a circuit to get an overview of circuit operation, a transmitter operator remotely selects which sample to monitor. Early remote control systems used the electromechanical stepper switch from the telephone industry to select among the metering samples. The step pulses for the stepper were generated with a mechanical telephone dial. More recent remote controls use individual mechanical relays (reed relays or other relays designed for dry switching) for analog multiplexing. The most recent remote control designs use integrated circuit multiplexers for analog multiplexing. These circuits act like a mechanical rotary switch. A typical multiplexer may accept three address inputs to determine which one of eight inputs will be connected to the output. The IC multiplexers cost less than relays, are faster, do not have a cycle limit, and take less space. They do, however, have a higher leakage current than relays and do not tolerate voltages higher than the supply voltage. If the voltage on an input is higher than the supply voltage, stray currents will generally develop in the circuit, causing erroneous readings on all inputs. With appropriate precautions, however, integrated circuit multiplexing is well suited to the job.

Telemetry Encoding

Once the analog multiplexer has selected a sample for transmission to the control point, it is necessary to encode the data into a form suitable for the communications medium involved. If the circuit has dc continuity, the output of the multiplexer can be applied directly to the circuit. If the circuit does not have dc continuity, further encoding is necessary.

Analog FM Encoding of Telemetry

A low cost method of converting a dc sample voltage to a form suitable for transmission over a voice grade or other ac circuit (such as subcarrier) is to use the sample voltage to drive a voltage-to-frequency converter (V/F) or frequency modulated oscillator. A block diagram for a system using this technique is shown in Figure 4.7-11. A zero sample voltage results in one frequency, full-scale results in another. A half-scale sample results in a frequency halfway between the two. One such system utilizes 800 Hz for zero sample voltage and 1200 Hz for full-scale. This is suitable for transmission over any voice grade circuit. A similar scheme uses 20 Hz to 30 Hz to transmit analog metering below program audio on AM stations or subcarriers. This system can also be used to generate higher frequencies for use as subcarriers. It is more common, however, to use the audible or subaudible metering as input to a subcarrier generator. This may result in FM (the remote control V/F converter) on FM (the subcarrier generator) on FM (an FM broadcast or TV aural transmitter). This approach is used because of the common voice grade interface.

At the control point, an F/V converter (or FM demodulator) develops a dc voltage that corresponds to the sample voltage. This voltage drives the meter or display at the control point. Accuracy of these systems is limited by the tracking between the V/F and F/V converters. Typical errors include zero drift, gain drift, and nonlinearity. The zero and gain errors can he minimized by frequently adjusting the F/V converter offset and gain while driving the V/F with zero and full scale samples.

Calibration of these systems typically involves a transmitter technician using the telephone to call the operator at the remote control point. The transmitter

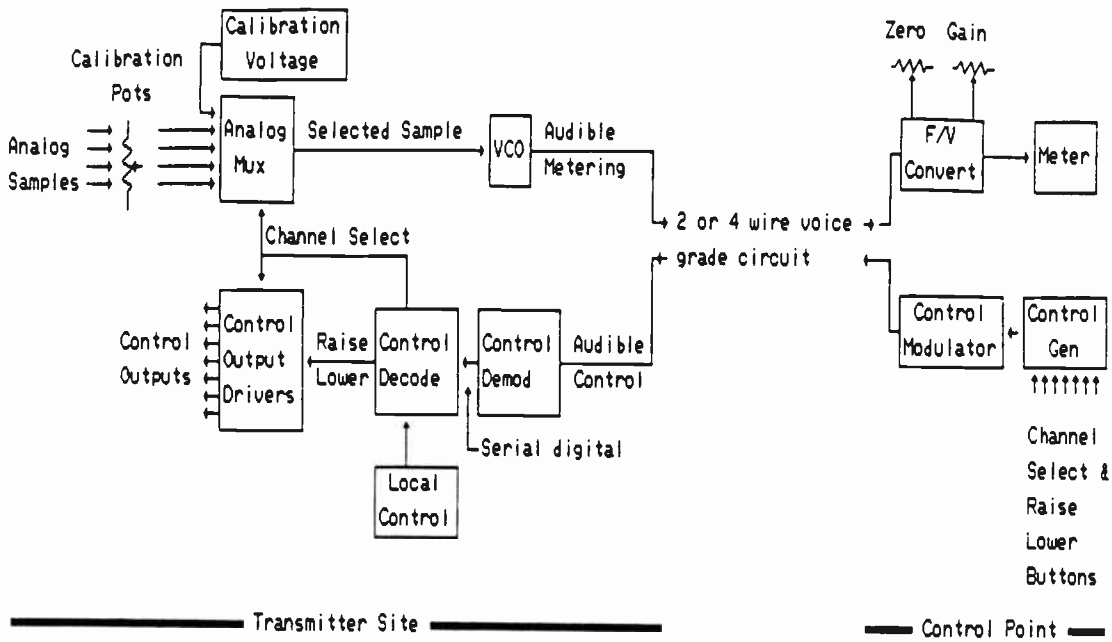


Figure 4.7-11. Typical analog remote control block diagram.

technician then adjusts calibration controls at the transmitter site until the operator at the control point says the reading matches the reading at the transmitter site.

A/D with FSK Transmission

Another method of transmitting the analog sample voltage to the control point consists of splitting a digital voltmeter into two parts (see Figure 4.7-12). The A/D converter is put at the transmitter site, and the display is put at the control point. This typically involves adding some data formatting circuitry to the output of the A/D converter and driving the transmit side of a universal asynchronous receiver transmitter (UART). The serial data out of the UART drives the transmit side of a modem. This modem typically uses frequency shift keying at 300 to 1,200 bps. The digital output of the A/D also drives a local display, allowing calibration at the transmitter site.

At the control point, the incoming audible FSK data is demodulated back to the original serial digital form. The serial digital data then drives the receive portion of a UART. The parallel output of the UART then drives data decoding and display circuitry. Since digital coding was used in transmitting the output of the A/D to the control point, the control point display will indicate the same as the transmitter site display. If there is an error in the received digital data (which is generally detected by a parity check on each byte received), the display is blanked and an error lamp is lit.

Control Encoding

As mentioned previously, if dc circuits are available, pulses of loop current or voltage may be used to send control to the transmitter site. Many recent systems use some form of tone encoding.

Frequency Shift Keyed Control

The ready availability of voice grade circuits has encouraged the design of systems utilizing audio tones for the transmission of control signals. The most common technique is the use of frequency shift keying (FSK), the same technique described in the A/D with FSK section. Early systems utilized a telephone dial or integrated circuits emulating a telephone dial to generate a string of pulses. The number of pulses in the string determined what channel was selected. More recent systems utilize binary coding. Since a UART normally sends eight data bits, six could be allocated to channel select, and two to control (one raise, one lower). This allows a simple 64 channel system to be built (since six bits can code decimal numbers 0 to 63). Since an error in a control byte transmitted can have serious consequences (turning transmitters on and off), many systems send the data at least twice and only accept it as error-free if the two transmissions agree.

Typical Nonprocessor-Based Design

Figure 4.7-11 shows the design of a typical analog remote control system. Digital circuitry is used to encode operator instructions through the channel select and raise/lower buttons. The serialized control data is sent to an FSK oscillator (the control modulator). This generates an audio tone at 400 Hz for one digital input level and 300 Hz for the other digital input level.

This audio control signal may be applied directly to a two-wire or four-wire telephone line (or other

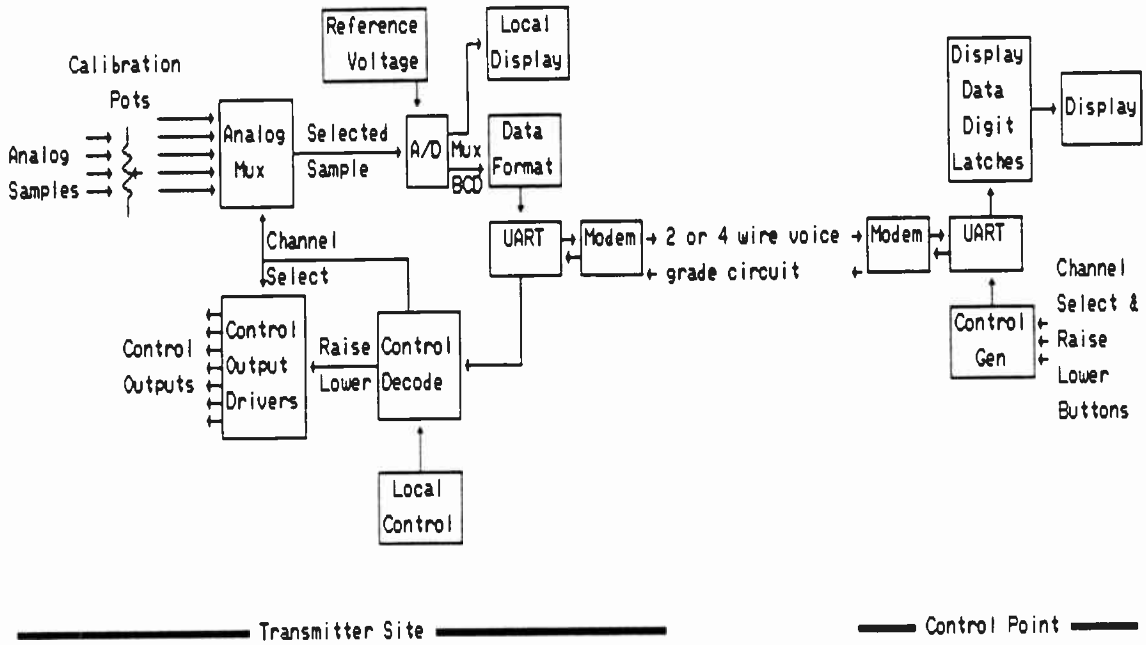


Figure 4.7-12. Typical digital remote control without processor.

audio circuit). It may also be further modulated onto a subcarrier for carriage above programming on an STL transmitter.

At the transmitter site, the FSK audio is demodulated back to the original serial control data stream. This drives the control decode circuitry, which determines which channel the operator selected and whether a raise or lower button was pushed. The channel number is sent to the analog multiplexer (which may be relays or an integrated circuit analog multiplexer) and the control output drivers (which may be transistors, integrated circuit output drivers, or relays). When a channel is selected and a raise or lower button is pressed, the selected channel control line is enabled.

Sample voltages from transmitters and other site equipment are scaled (multiplied by a constant) by *calibration pots*. During system calibration, these are adjusted so the control point meter reads the same as the meters at the transmitter site. The outputs of the calibration pots drive the analog multiplexer. This sends the sample voltage selected by the operator to the voltage controlled oscillator (VCO). Note that a calibration voltage is also shown. Such systems will typically allocate one or two channels of the analog multiplexer to a zero and full-scale (occasionally half-scale) voltage. These reference voltages allow the operator at the control point to adjust the F/V converter offset (zero) and gain controls to compensate for drift in the VCO and F/V.

The output of the analog multiplexer drives a VCO. The VCO will typically generate 800 Hz for a zero reading and 1,200 Hz for full-scale. If subaudible metering is used, 20 Hz represents zero and 30 Hz full-scale. The 800 Hz–1200 Hz metering frequency range is well separated from the 300 Hz–400 Hz control frequency range, allowing both signals to be put on a single two wire telephone line, yet be easily separated at each end using filters. This combining of two signals onto a single circuit by using different frequencies is a form of frequency division multiplexing.

Finally, the 800 Hz–1,200 Hz metering signal gets back to the F/V converter at the control point. The dc output voltage (proportional to the deviation from the nominal frequency) drives a front panel meter, which the operator reads. By selecting each channel and noting the meter indication, the operator can monitor several transmitter site parameters. Further, by selecting a channel and using the raise and lower controls, various transmitter site controls may be activated.

Nonprocessor Digital Remote Control

Figure 4.7-12 shows the design of a typical digital remote control that does not use a microprocessor. It is quite similar to the analog system described previously.

The calibration pots, analog multiplexer and control output drivers are the same as the analog system. The digital system does not, however, require calibration voltages to be sent to the analog multiplexer. The A/D converter typically accepts a reference voltage (its output is the ratio of the unknown voltage to the reference voltage). The A/D automatically calibrates and zeroes itself during each conversion.

The output of the A/D in most such systems is multiplexed *binary coded decimal* (BCD). The decimal code for each digit is represented as a 4 bit number on four wires. The code for each digit is multiplexed onto these same four wires. Digit select lines out of the A/D tell the outside circuitry which digit's data is present at any particular time. This data may be latched and decoded to drive LCD or LED displays. It may also be merely decoded (not latched) to drive multiplexed LED displays. The output of the A/D drives a display at the transmitter site, allowing the calibration pots to be adjusted at the transmitter site, since the transmitter site display will exactly track the control point display.

The output of the A/D also drives data formatting circuitry which converts the multiplexed BCD to something suitable for the UART. The transmit side of the UART accepts up to 8 bits of parallel data and a strobe pulse. It then serially sends the data bits (along with some synchronizing bits). Typical data formatting circuitry selectively gates the digit select outputs of the A/D to the transmit strobe input of the UART, causing each digit of BCD data to be transmitted. The multiplexed BCD is connected directly to four input lines of the UART. The digit select gating allows the A/D data to be sent to the UART without latching the multiplexed BCD data out to parallel BCD (which would take lots of parts). Typically, a couple more bits of the UART are used to transmit a *digit identification code,* corresponding to the digit being transmitted. At the receive end, the digit identification code is used to route the incoming BCD data to the proper digit of the display.

The output of the UART is serial digital data. This data is sent to a modem for transmission over a voice grade circuit. Often, Bell 103 modem tones are used. These modem tones allow the transmission of 300 bps in each direction over a single two wire voice grade circuit (again, using frequency division multiplexing). Bell 103 uses FSK with 1,270 Hz representing a 1 and 1,070 Hz representing a 0 in one direction. Data going the other direction uses 2,225 Hz for 1, 2025 Hz for 0.

The audible metering data is received by the modem at the control point. It is converted back to serial digital data by the receive side of the modem. The UART changes the serial digital data back to the original parallel form (up to 8 data bits plus a strobe). If the previously described encoding scheme was used (4 bits of BCD, 2 bits of digit select), the digit select code bits may be used to drive a demultiplexer integrated circuit. This *demux* routes the receive data strobe from the UART to the clock input of a latched display driver for each digit. The latched display drivers for each digit are being fed the received BCD data from the UART, but only capture the data for the appropriate digit when the strobe for that digit is presented.

In a similar manner, front panel controls may be sent to the transmit side of the UART, where the data is serialized for transmission to the transmitter site. The transmit side of the control point modem converts the data to a form suitable for transmission over a voice grade circuit. At the transmitter site, the receive side of the modem converts the control data back to the original serial digital form. The receive side of the UART converts the control data back to the original parallel form.

Microprocessor Based Remote Control

Each of the techniques described so far has used standard nonprogramable integrated circuits to build a system dedicated to transmitter control. Each system was a digital voltmeter with a remote input select switch and display

Microprocessors greatly enhanced the capabilities of transmitter remote control systems. In microprocessor based systems the basic components of the remote control remain the same. There is still an analog multiplexer, an A/D converter and a modem. Instead of being connected in a manner fixed in hardware, they are all connected as input/output (I/O) devices on a computer bus. The computer processor then addresses each of these devices, telling them what to do and picking up data. The complexity of the system is moved from hardware to software. The use of computer techniques allows features to be added to the system at very low production cost (as compared to providing the same features with dedicated hardware).

Figure 4.7-13 shows a block diagram of a rather large microprocessor based remote control transmitter site unit. Most systems will not contain all of these elements, but processor based systems can be designed to allow for substantial expansion.

Processor

The *processor* executes a computer program. It reads instructions and data from the system memory. It can read data from input ports and write data to output ports (commonly combined into I/O ports). It can make calculations and decisions based on those calculations. It can also write the results of calculations back out to memory or I/O.

The processor has three buses to communicate with the rest of the system—the *control bus, address bus* and *data bus.* In simple systems (such as a typical remote control), the address bus is driven only by the processor. The processor uses this bus to tell memory and I/O devices what location is to be read or written. In transmitter remote control equipment, the address bus is normally between 11 and 20 bits wide.

The data bus is a bidirectional bus. It is driven by the processor when the processor is writing (sending) data to a memory location or an I/O device. It is driven by the memory or I/O device when the processor is reading data. The data bus in transmitter remote control equipment is generally 8 or 16 bits wide. The processor uses the address bus and the control bus to tell other devices on the data bus whether data is for that particular device, and where in that device the data is supposed to go or be read from.

The control bus varies between different processor types. In Motorola style processors, a +read/−write line has the same timing as the address lines. It is high during a processor read and low during a processor write. These processors then have a common (between read and write) strobe line (called E or phase 2). This

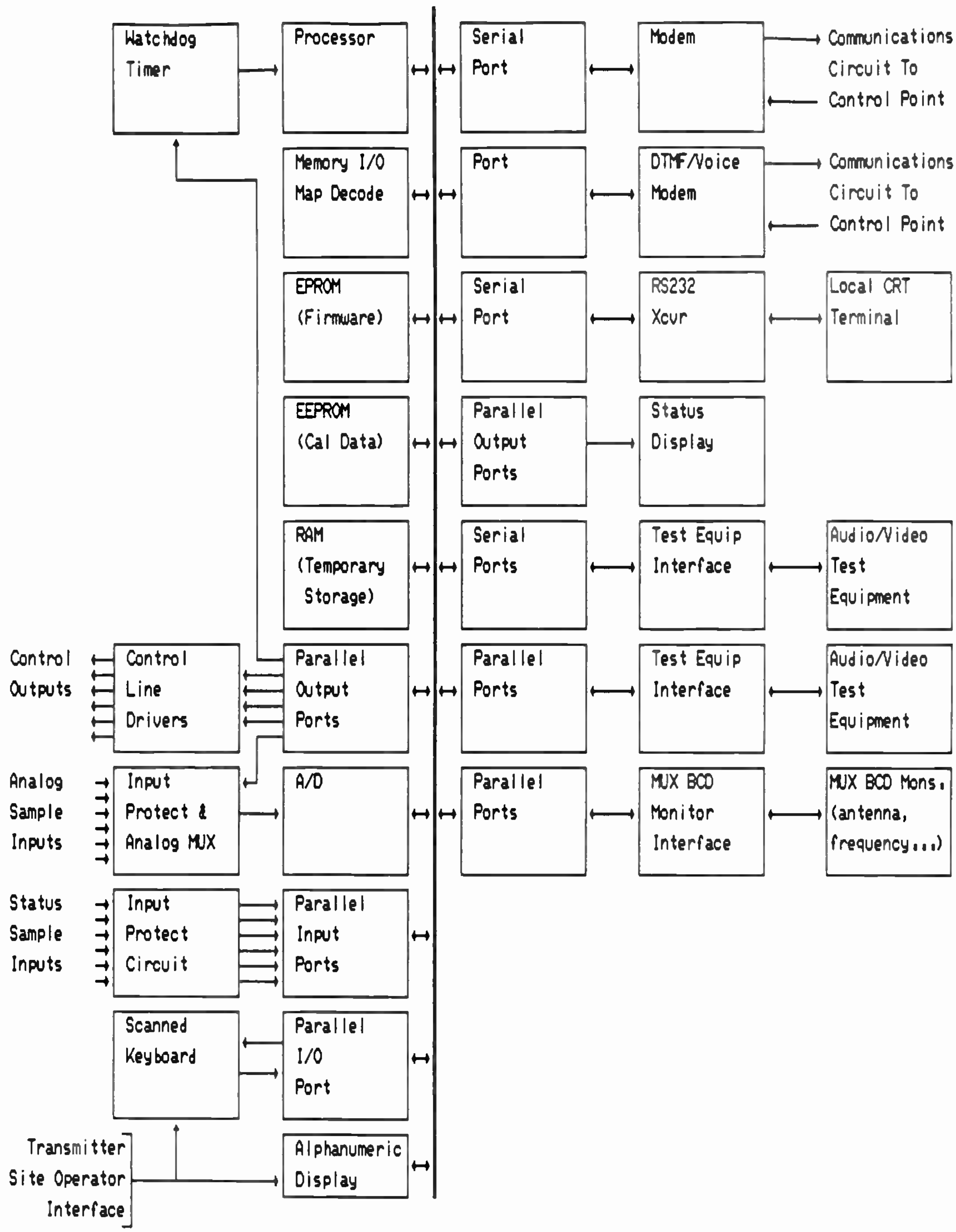


Figure 4.7-13. Transmitter site unit of microprocessor-based remote control.

line goes high during a read or write. Data is enabled on the leading edge of the strobe and latched and disabled on the trailing edge. Intel style processors use two separate strobes, one for read and one for write. These have about the same timing as the common strobe in Motorola style processors.

Another variation in the control buses deals with how I/O devices are handled. Motorola style processors have no I/O instructions. Instead, memory instructions are used for I/O. If the address that data is sent to has an I/O device instead of a memory device, the data is output. The processor does not know where the data is sent. It just puts an address on the address bus, puts the data on the data bus, then strobes the enable line. What device (if any) picks up the data is not important to the processor. Similarly, the processor can read data from either memory or I/O. Motorola style processors use the same instructions for both. This is called *memory mapped I/O.*

Intel style processors have a *separate I/O map*. Separate I/O instructions setup the address bus and put data on the data bus, just as memory write instructions. However, an I/O write strobes the I/O write strobe instead of the memory write strobe, allowing I/O data to go to the I/O device instead of a memory device, which may be at the same address. Systems based on Intel style processors may also use memory mapped I/O.

The other main control bus lines are the *interrupt lines*. When an I/O device enables an interrupt line, the processor stores away its current state, then executes an *interrupt service routine* (ISR). The number of interrupt lines varies from processor to processor. When a particular interrupt line is enabled, the processor vectors to a particular address to find the ISR for that device. Several I/O devices may share the same interrupt line. When this is the case, the ISR polls each device using that interrupt line to determine which one generated the interrupt. On completion of the ISR, the processor executes a *return from interrupt* instruction, which restores the processor state to what it was prior to the interrupt. The main program continues on as though nothing happened. Interrupts are commonly used to allow a processor to appear to multitask.

Watchdog Timer

Occasionally, something will go wrong in a computer system. When this occurs, the processor must be reset. The reset loads the program counter (or instruction pointer) with the beginning address of the system program, allowing it to start over. System programs typically include an instruction to set a bit of an output port high at one point in the program, then set it low at another point. If this fails to occur within some time period, the *watchdog timer* times out resetting the system. This ensures that a system crash will not disable the system.

Memory-I/O Map Decode

This portion of the remote control may be done in one chip, or may be spread over several chips. The map decoder recognizes when particular addresses are present on the address bus, then generates a *chip select* for the device that is to be selected.

Erasable Programmable Read Only Memory

The *erasable programmable read only memory* (EPROM) holds the program the system executes when the power is first applied. Depending upon the system, this may be the only program the system ever executes. Some systems may hold additional software in EEPROM or RAM. The program that tells the system how to be a remote broadcast transmitter controller (instead of, say, a microwave oven controller) is held in this EPROM. The program is just a series of binary numbers. Under direction of the processor, the EPROM puts the contents of its various memory locations (more than 1,000,000 8 bit locations are available in today's EPROMS) on the processor bus. The processor reads these instructions, then executes them.

The program in the EPROM is often called *firmware*. The firmware may be written in one or more of several computer languages. In each case, a program (either an assembler or a compiler) converts the program into a series of binary numbers. These binary numbers are programmed into the system EPROM. The processor reads the program from the EPROM and executes the program.

Electrically Erasable Programmable Read Only Memory

Most systems now include *electrically erasable programmable read only* memory (EEPROM) or battery backed RAM to hold system setup information. Some systems may also run firmware from EEPROM, allowing field upgrade of the firmware without opening the box. Such upgrades can be accomplished by downloading from an external computer, an internal floppy drive or through the dial-up modem.

EEPROM is similar to EPROM in that it is nonvolatile (it retains the stored data when power is lost). It is similar to RAM since it can be reprogrammed in circuit (EPROM is typically removed from the circuit, erased with ultraviolet light, reprogrammed, then put back in the circuit). Most recent EEPROMs use the same bus interface as static RAMs, easing system design considerably. EEPROMs can be written to, the same as RAM, except that a time delay of 1 msec to 10 msec is required before another read or write to the chip. The system firmware must take the EEPROM write delay into account (it cannot stuff calibration data into the chip at a typical processor speed of a byte per microsecond or faster).

The ease of programmability of system EEPROM has its drawbacks. If there is a processor crash, it is easy for the processor to write garbage into the EEPROM, causing loss of calibration data. Most systems, therefore, include extra write protect hardware for the EEPROM. This hardware typically consists of an output port that drives gating circuitry for the write to the EEPROM. Prior to writing data to the EEPROM, the processor must write a specific code to the write protect port. After the data has been sent to the EEPROM, the processor writes another code to the write protect port, disabling further write strobes from reaching the EEPROM. It is unlikely that a crashed processor would write the proper code to the write protect port, thus protecting the EEPROM data until a watchdog timeout resets the system. The only other limitation on EEPROM is the limited number of write cycles. Each byte of EEPROM may typically be written 10,000 to 100,000 times. Since calibration and setup of transmitter control systems is typically done no more than once a week, this limitation is generally not serious.

Some systems use battery backed CMOS RAM to hold system calibration data. This has several advantages over EEPROM. The RAM can operate at full processor speed (writes at more than one byte per microsecond) and there is no limit on the number of write cycles. The RAM is normally battery backed with a lithium battery, which has an expected shelf life of ten years. The addition of the battery and other power switching circuitry makes the system design more complex, though some companies sell battery backed RAM subsystems (the memory, battery and all support circuitry contained in a standard 24 pin package, the same package used by RAM or EEPROM). One battery backed RAM subsystem also includes a battery backed real time clock and all associated circuitry, again in the 24 pin package. Use of this chip further simplifies system design, when a real time clock is required. The battery backed RAM suffers the same too easy to write problem as EEPROM. Many systems, therefore, include write protect circuitry for the battery backed RAM.

The use of nonvolatile memory for system calibration and setup has vastly increased the capabilities of transmitter control systems. A calibration pot may be replaced with 4 bytes of EEPROM. A single byte of EEPROM can replace 8 DIP switches.

EEPROMS (or battery backed RAM) in transmitter control systems typically include 2 kB to 32 kB of nonvolatile memory (where 1 kB is 1,024 bytes). The data included in the nonvolatile memory may include the following for each analog channel:

- A scaling factor, replacing the calibration pot and decimal point positioning logic.
- An alphanumeric label identifying the parameter (such as ICP for common point current).
- An alphanumeric units identifier (such as kV for kiloV).
- A curve or formula identifier (such as linear or square law).
- A *sample delay* to allow for the settling time of external monitoring equipment (such as AM antenna monitors).
- Alarm and action limits on the parameter.
- Alarm and action codes for each limit. These may adjust transmitter power on one set of limits and alarm on another set of limits.

For each status input:

- Active high or low (light indicator when input high or low).
- Latch or momentary. It may be desirable to have some status indicators remain on even after the status input has gone back to the false state. The status indicators are then reset by operator command.
- Status change alarm or action code. A status change may cause some control command to be issued or enable an audible alarm.

Other site specific data:

- Communications parameters for intersite communications (typically between the transmitter and the transmitter control point).
- Site identification number. Many systems use the same firmware in control point and transmitter site equipment. The site identification number tells the processor which portions of the firmware apply to this site. This site number is also used to recognize messages intended for this site in multisite systems.
- In multisite systems, site intercommunication data such as site access slot timing, highest site number, and site message routing tables. Communications parameters for any other serial devices (such as CRT terminals, test equipment, video monitors, etc.).
- Coding for interface to specific monitors at the site (such as those connected to the remote control through RS232 or other serial links, IEEE488 or other parallel links, and multiplexed BCD).
- Time specific data, such as generating control pulses at specific times (perhaps AM pattern change), or changing limit tables based on time (different limits used day or night).

The variety of data held in the system nonvolatile memory is limited only by the system firmware. The use of specific locations can be changed through system firmware updates (EPROM swaps), allowing features to be added later with no hardware changes. Finally, the packages and pinouts for all byte-wide memory devices have been standardized. This allows more memory to be added to the system (through addition of chips or the substitution of larger chips in the same sockets) or the memory balance (between EPROM, RAM, and nonvolatile memory) to be adjusted as the system design advances. This is a vast improvement over calibration pots and DIP switches!

Random Access Memory

Random access memory (RAM) is used for temporary storage in the processor based system. The two major RAM types are *static* and *dynamic*. Static memory remembers data as long as the power is applied. Dynamic memory remembers data for about 2 msec. After 2 msec, dynamic memory must be refreshed. Due to the simplicity of design of systems utilizing static memory, it is commonly used where 64 kB of RAM or less is required. In larger systems, the added expense of refresh circuitry is offset by the lower cost of dynamic RAM.

The processor uses RAM to hold *return addresses* so it can find its way back after a subroutine call or interrupt. RAM is also used to form buffers for I/O devices. With appropriate firmware, the system can send large blocks of data (such as a full screen of characters) to a *device driver*. This firmware then holds the data in RAM and sends it to the output device (such as a serial port or modem) one byte at a time, as the port is ready to receive more data.

RAM buffers may also be used in receiving data from an input port. This allows more flexibility in the system firmware, since it need not continuously poll a specific input device. Without an input buffer or continuous polling, some input data could be lost (the firmware may be dealing with the A/D converter when data arrives from the modem, causing modem data to be lost). The RAM input buffer holds received data until the main firmware is ready to deal with it. Another RAM input buffer is used with the communications port that receives data from the control point.

To ensure data is received error-free, it is typically sent in *packets*. Each packet of data includes error detecting data at the end of the packet. The packet cannot be dealt with until the entire packet is received. The RAM buffer holds received packets until they are complete. The system firmware then determines if the packet is error-free. If so, the packet may be acknowledged (depending on system design) and passed on to other sites or firmware at this site. If a packet has an error, a negative acknowledgement is typically passed back to the originating site.

The amazing flexibility of processor based systems is again evident in the uses of system RAM. All of these uses (and more) may reside in the same chip. The allocation of that memory may change under program control and with system firmware upgrades (EPROM swap). In a pure hardware (nonprogrammed) system, each buffer would probably be implemented using another chip, considerably adding to hardware complexity.

Parallel Output Ports

A parallel output port may consist of an 8 bit latch. When the processor does a write to the appropriate address, the memory I/O map decoder clocks the latch, causing it to capture the data then on the data bus. This data is held until the processor does another write to the same address. As can be seen from Figure 4.7-13, parallel output ports serve several purposes. Whenever an operator needs to generate a logic level under program control, a parallel output port is used.

One bit of a parallel output port is used to drive the watchdog timer. Additional lines of a parallel output port (not shown) drive the EEPROM write protect circuit.

Additional port outputs drive the control line drivers. The output port itself is typically a low power Schottky transistor logic (LS) or high speed complementary metal oxide semiconductor (HC) compatible device, which has limited driving capabilities. The port outputs drive control line drivers, which may be discrete transistors or peripheral driver chips. These open collector control outputs may be sent directly to the rear panel (through RF filters) to drive transmitter site equipment, or may drive interface relays within or outside the remote control unit. Under firmware control, these control lines may be set high or low or pulsed high or low. How these lines are pulsed or set is typically a combination of user action (command keystrokes), system parameters (automatic control) and user data in the EEPROM. These lines replace the raise and lower control lines of nonprocessor based systems. For convenience, many processor based systems still refer to the control lines by a raise/lower designation.

Additional parallel output ports may be used to drive status indicators (typically LEDs). These may be driven directly by the output port or may be driven by a peripheral driver chip. The various LEDs can be lit or extinguished under processor control. These may reflect the status inputs at this site or another site, or other on/off conditions.

Additional parallel output port lines drive the address lines of the analog multiplexer. These determine which of the many analog samples are routed to the A/D converter.

A/D Converter and Analog Multiplexer

The A/D converter serves the same purpose as that in the nonprocessor based digital remote control. It converts the analog sample voltage from the analog multiplexer to a digital code. Of the several A/D techniques available, transmitter control systems typically use integrating (dual slope or residue multiplication) or successive approximation. For the same cost, an integrating A/D offers higher resolution (12 bits versus 8) while the successive approximation A/D is faster (50,000 conversions per second versus 20). Although the same A/D used in the nonprocessor based system could be used with a processor, A/Ds are available that are designed to interface directly with processor buses. The processor may read the A/D to see if a conversion is complete and to pick up the converted data. It may write to the A/D to start a conversion, select an input range, select ac or dc or select an input (some include analog multiplexers).

The A/D is driven by the analog multiplexer. Any of the previously described analog multiplexing techniques may be used. Further, the sample selected by the multiplexer may be directly selected by the operator (who may be interested in only a particular channel at the moment), or may be scanned through all the inputs that are used. When the multiplexer is scanned, the resulting conversions are usually held in RAM, making them immediately available to the firmware and the operator.

Solid-state analog multiplexers typically will not accept voltages higher than the supply voltages (either positive or negative). Many multiplexers are protected from damage due to higher than supply input; however, they will typically have excessive leakage into other channels under this condition, causing erroneous readings. The input protect circuitry typically protects the multiplexer from damaging voltages and those voltages that may cause improper operation.

Scanned Keyboard Port

A transmitter site operator needs to interact with the system for setup and calibration. The local display and keyboard serve this purpose. With the addition of four pullup resistors, a single parallel I/O port (4 bits input, 4 output) can be used to scan a 16-key keyboard (typically arranged as four rows and four columns).

Alphanumeric Display

This display is used by the processor to talk to the operator. In simpler systems, the display may only display numeric data. In more complex systems, the display typically displays 32 alphanumeric characters. The use of a full alphanumeric display allows the processor to explain the operation of the system. Each key may have several uses (reducing the keyboard hardware required). The display allows the processor to inform the operator the meaning of each key at this particular point in the operation of the system.

Many systems use alphanumeric display modules that interface directly to the processor bus. A write by the processor to the display may put a character on the display, position the cursor, clear the display or various other control functions. The processor can also read data from the display module. It can typically read a status bit indicating whether the display controller is ready for another character (the display controller is generally slower than the system processor), read the cursor position and various other parameters of the display controller.

Alternate Display Techniques

When large amounts of information are to be displayed, it is difficult to beat the economics of a CRT display. Transmitter and control point operator interaction with the system can be accomplished with a full alphanumeric keyboard and CRT at the site.

Alternate Interaction Techniques

Some remote control systems do all operator interaction (including transmitter site setup and calibration) without a display. All operator interaction with the system is through dual tone multifrequency (DTMF) or Touch-Tone and voice. A standard telephone replaces the keyboard and display previously described. The processor either synthesizes voice in software or sends recorded speech to a (relatively) high speed D/A converter. The output of the D/A is sent to the local (or a distant) telephone. The operator responds to inquiries from the system using the DTMF keyboard on the telephone. This technique has proved quite popular and has been awarded an U.S. patent. Again, however, when large amounts of data are to be displayed, it is hard to top a CRT display. Most systems with DTMF/Voice also offer serial ASCII interfaces (to drive a CRT) as an option.

Digital Monitor Interface

Not all transmitter site parameters monitored by the remote control system can be easily changed to dc voltages. Some monitors digitally derive a parameter, then convert it to an analog sample for interface to the remote control. This is a rather indirect method of getting data from one instrument to another (digital data converted to an analog voltage, then digitized again). A direct digital connection would be simpler and more error-free (no D/A, A/D tracking problems). Several methods exist for handling the digital interface. These can be divided into serial and parallel interfaces.

Serial Monitor Interface

Various serial interface standards exist. The most common is EIA-232-D (RS232) with asynchronous data. Typically, not all the EIA-232-D lines are used. The two signal lines most commonly used are *transmit serial data* and *receive serial data*. The remote control processor can send requests to the external monitor by writing to the appropriate serial port in the remote control. The monitor sends data back to the remote control through the same serial port. The processor can read this data from the same serial port.

More and more transmitter site monitoring equipment is equipped with an EIA-232-D port. That equipment is often designed to be directly connected to a modem, allowing remote access to the equipment. If each piece of equipment got its own phone line, a transmitter site could end up with many dial-up lines. A *mini private branch exchange* (PBX) is now available to serve this need. An operator can call the site, key in an extension number and be connected to the appropriate piece of equipment (whether it be the transmitter control, the modulation monitor, a video signal analyzer etc.). A similar RS232 switch is available. It would be placed between the modem and the transmitter site equipment. Finally, transmitter remote controls may serve as multiplexers or routers for RS232 data from transmitter site test equipment.

EIA-232-D allows one port to talk to one device. It also uses single ended data transmission (one signal line fed against ground). EIA-422-A (an electrical specification that is part of EIA-449), however, uses balanced lines (two wires driven differentially). Some EIA-422-A applications operate in a *multidrop mode*, allowing several devices to share the same port. When operating in a multidrop mode, each device connected to the port is addressed by the processor (polled) when the processor needs data from that device. This technique allows several devices to share a single serial port, saving on hardware.

Parallel Monitor Interface

Various parallel interfaces exist. For example, a frequency monitor may present its data in parallel BCD, where each digit of its indication is transmitted over four wires. Obviously, this requires a lot of wires and a lot of interface hardware.

An antenna monitor may provide multiplexed BCD data. Four wires carry all the BCD data and additional *digit select* lines identify the digits. An antenna monitor will also require various control signals, such as tower selects. Again, a lot of wires and interface hardware are required.

Further complicating the BCD interface is the wide variety of software that would be required to deal with each instrument. Hewlett Packard recognized the problem with existing test equipment interfaces and invented the Hewlett Packard Interface Bus (HPIB). This eventually became the IEEE488 bus, also known as the general purpose interface bus (GPIB). The IEEE488 bus standard includes well-defined hardware specifications and addressing specifications (it is a multidrop parallel bus). Some standards also exist for command and response messages. At this writing, the IEEE488 bus is not widely used in broadcast remote control applications.

Intersite Communications

Most transmitter remote control systems utilize low speed (300–1,200 bps) asynchronous data to communicate between the control point and the transmitter. In Figure 4.7-13, this is shown as a serial port driving a modem. Some systems offer a variety of modems that plug into the system. One modem may be optimized for two-wire full duplex voice grade circuits. Another may be optimized for four-wire voice grade circuits. Another may directly generate signals suitable for use as subcarriers on STL and broadcast transmitters. It is also possible to bring out a serial digital interface to communicate with the control point. The station can then transmit the data digitally to the control point. This digital transmission may involve the use of external modems, a fiber optic link or multiplexing the data into a higher speed digital link (such as a digital STL).

The data that is sent to the control point must, of course, be of a form that the control point equipment can handle. This data may be ASCII data ready for immediate screen display (directly driving a CRT terminal) or a stream of site data where the control point equipment does error checking and extracts the required data for display. In each case, the data and its encoding (modem type) is optimized for the control point equipment and the communications channel. One widely available piece or control point equipment is the standard DTMF telephone. Since this does not have a display, data is sent to it using voice synthesis. Commands from the phone are sent to the transmitter site using the DTMF tones. The voice synthesizer/DTMF detector can be thought of as another modem that takes digital data and formats it for transmission over a voice grade circuit.

Summarizing Transmitter Site Hardware

Figure 4.7-13 shows many of the elements that *may* be present in the transmitter site unit of a transmitter control system. Some systems may have more I/O devices hung on the bus, others may have less. There are microcontrollers available that combine many of the hardware requirements for such a system on one chip. Some microcontrollers contain a processor, EPROM, EEPROM, A/D converter, analog multiplexer, watchdog timer, serial port and parallel I/O on one chip. If one of these chips can meet system requirements, a very low chip count system can be built. However, once the design does not fit on a single chip, the system design may become more complex using microcontrollers in their expanded mode as opposed to using a standard processor.

Control Point Hardware Design

The *control point equipment* is used to display data gathered by the transmitter site equipment (or announce it, in a voice/DTMF system), and transmit operator commands back to the transmitter site. In most cases (again, excepting voice/DTMF systems), the control point hardware is another microprocessor based system using many of the same components as the transmitter site equipment. It has been fairly standard practice to use the same hardware for control point and transmitter site equipment, generally eliminating the analog inputs and control outputs and some other special I/O at the control point. The control point equipment would include the processor, memory, a display, a keyboard, status display and modem (for communicating with the transmitter site).

If the system were to drive a control point CRT terminal, typically the intelligence for driving the terminal would be located in the control point end of the remote control system. Since the CRT terminal is also a microprocessor based system that displays received data and transmits operator keystrokes, it appears possible to replace the control point end of the system with a standard CRT terminal and a modem. This is the basis of some more recent remote control systems. The use of a standard CRT terminal as the control point equipment reduces the system cost. However, most standard CRT terminals do not include a watchdog timer, allowing the terminal processor to crash in a power dip. An external watchdog timer may be added, if desired.

Use of Standard Hardware

Broadcast transmitter control equipment has typically been designed specifically for the industry. Standard components are used, but the circuit boards are designed by the equipment manufacturer. Computer hardware, however, has become somewhat generic. In many cases, it is more cost effective for a manufacturer to buy standard computer hardware (that uses either the Industry Standard Architecture bus or another industrial computer bus). In such cases, the manufacturer may buy the processor board, video boards, serial and parallel I/O boards, modems, disk interfaces, etc. While standard A/D, status in and control output boards exist, the required additional RF filtering required for operation in a broadcast environment normally makes it more cost effective for the equipment manufacturer to design and build this board. The distinction between different systems becomes a matter of which standard hardware has been plugged into a standard system board along with the unique software for each system. Product differentiation has moved from hardware to software.

Lower cost systems, especially those based on a microcontroller with the majority of the hardware on one chip, continue to be based on custom hardware.

System Firmware Considerations

The firmware design of a minimum system (one emulating the nonprocessor based system of Figure 4.7-12) can be implemented relatively easily. However, in an effort to reduce parts count (and manufacturing costs), there is a trend towards moving as much of the design as possible from hardware to software.

For example, the calibration pots are replaced with scaling factors and multiplication routines. Analog multipliers (power-to-linear converters) are replaced with software. Once the system includes a processor, powerful features can be added with minimal manufacturing cost (though the design cost may be the same or higher than designing a similar feature in hardware). These added features may include automatic limit checking, transmitter adjustment, transmitter logging, full screen display, etc.

Intersite Communications

In a simple system with a single control point and a single transmitter site, data sent by the transmitter site to the control point is displayed and operator keystrokes are sent to the transmitter site for appropriate action. These communications are normally carried out using simple data packet techniques so that any erroneous data is ignored. If an error-free data link can be established, the display (perhaps a CRT terminal) can be connected directly to the link. Such an error-free link can be established by moving the error correction responsibility from the control point box to the modem. All modems currently on the market include error correction firmware. Even a noisy line should not cause erroneous data.

It gets more interesting when there are multiple transmitter sites. If dial-up lines are used, the telephone company handles all the site selection and data contention problems. DTMF/voice systems may use a single telephone with an automatic dialer. If a site has a problem to report, it calls it in. If the operator wants to poll a site, a single key on the dialer is pressed, and the operator is shortly in communication with that site.

A similar approach may be used with CRT terminal based dial-up systems. In such systems, programmable function keys on the terminal may be used to generate modem dialing codes for each of the transmitter sites. If a transmitter site calls in with an alarm report, the modem answers and puts the site information on the screen. The dial-up phone system is handling all the data routing problems. Further, only the data from one site at a time is displayed. If the CRT terminal is replaced with a computer and a modem, it can gather the last reported data from each site and put together a display showing a summary of the site information.

If dedicated circuits are used to interconnect the sites, it is the remote control's responsibility to route the data to the appropriate site. Two major methods exist. The first utilizes half duplex modems and is similar in design to local area networks (LANs) and packet radio data transmission. The second approach uses full time digital links (which may include modems to allow use of voice grade or other analog circuits) and is more similar to the switched telephone network or wire line (or fiber optic) packet data networks.

Multisite Data Packet Communications

In multisite communications, a single link is used to carry messages that may originate from any of several sites and be sent to any of several sites. To keep the messages straight, each one is put in a packet that includes the number of the originating site and the destination site. This information is included in the message header. The header consists of a flag identifying the beginning of a packet, the address information, and a message length count. A flag consists of a sequence of bits or bytes that is unique and guaranteed to not occur in the data stream (which would cause a false begin of header detection).

The structure of the header varies from system to system, but will often include the destination address (site number), the return address (site number), and a byte count for the message being sent in the packet. The destination address is used by each site in determining whether to accept a message, route the message to another site or ignore the message. The return address is used by the destination site to send back any requested data, a message acknowledgement, or a negative acknowledgement. The message byte count is used to determine where the end of the packet is. An alternative method is to use another flag sequence to identify the end of the packet. The header may also include a sequence number. As each packet is transmitted, the sequence number is incremented (and rolls over, back to zero). The receive site can then acknowledge a whole sequence of packets instead of each one individually. This generally speeds up the data transmission (a bunch of packets can be sent in one burst instead of sending them one at a time and waiting for acknowledgements). The actual contents of the data field of the packet may have varying meaning. Often one byte of the packet will identify the type of data contained in the packet. This packet type byte is then used to identify the contents of various fields of the data packet. The data packet ends with a frame check sequence (FCS). This is generally an 8 bit or 16 bit error detecting sequence, often a checksum, cyclic redundancy check or some similar error detection system.

Multisite Half Duplex Communications

Half duplex multisite systems have a single analog communications link between all sites in the system, so that all sites can hear all other sites. A message transmitted by any site is heard by all other sites in the system. This is similar to many LAN and packet radio systems. Each site receives all the packets that are exchanged in the system, but ignores all packets except those where the destination address matches the site number of that site. This is a fast communications system, since the message immediately goes to the desired site (and all others). Some method of keeping the sites from interfering with each other is required. Each site operates half duplex, keeping its modem carrier off line unless it has some data to transmit. Each site is continuously listening to the line. As such, it is somewhat similar to a tristate bus in a computer system (each device listens to the bus, but only the addressed device puts data on the bus). A variety of techniques are available to coordinate the data transmission on the channel. These include *master/slave, token passing, aloha, carrier sense multiple access* *with collision detection* (CSMA/CD) and *minislotted access.*

Multisite Full Duplex Communications

Another approach to routing data packets among several sites in a system is based on a continuous two-way data link linking each site to at least one other site in the system (a mesh topology). When a packet is received from any of the incoming links, the packet is first checked for errors. If the packet is error-free, the destination address is examined. If the destination address matches this site, the message is passed to the remainder of the firmware at this site (perhaps requesting a parameter from the A/D handler, etc.).

If the destination address does not match this site, a routing table (typically in EEPROM) is consulted. Based on the destination address, the routing table information and circuit loading, the received packet is placed in the output buffer for a particular output port, which drives a communications link to another site in the system. The routing table holds one or more port numbers for each site in the system. Where more than one port number is assigned, the system makes a decision as to which port is to be used based on system traffic. If the first choice port (the most direct path) has a lot of data in its output buffer, the second choice port may be used, spreading the traffic for optimum transmission speed (is it faster to use the indirect route or get in line for the busy port?).

On receiving a packet, a site does a similar evaluation. Note that a particular communications link will be carrying data destined for many different sites. A properly received packet that needs to be relayed to another site is put into a FIFO buffer for transmission to that site. Since each message must be fully received (for error detection) before it is passed on, the delay in getting a message from one end of the system will be more than in the previously described half duplex systems. This delay can be reduced by using smaller packets, though the header overhead then reduces overall system efficiency. This routed system does result in higher overall throughput since the system may carry more than one message at a time, while half duplex systems carry only one message at a time.

Further, there are no delays due to carrier bring up, shut down and carrier sense delays. Each site acts as a data switch and multiplexer along with its normal data acquisition and control activities. Each message is fully demodulated, error checked, retimed and remodulated when sent on to another site. This digital repeating prevents the accumulation of noise, distortion, and errors in the data.

Internet Protocol

The Internet Protocol (IP) defines a structure for data packets to be sent over a variety of circuits from any point (or site) to any other point (or site). At this writing, the IP is not widely used in broadcast transmitter control. However, use of this protocol would allow use of mass market hardware and software in broadcast transmitter control. The IP can be used on circuits owned by the broadcaster (such as subcarriers or other data channels on an STL) or dedicated circuits from a telecommunications carrier. The transmitter control data can stay entirely within a private network or can be introduced to the Internet, allowing control and monitoring of the transmitter from anywhere the Internet is available. In such instances, use of a secure IP is, of course, suggested. In such systems, the broadcast transmitter control unit becomes an *embedded server,* responding to requests from client units elsewhere on the network, just as other Internet servers respond to client requests.

Parameter Limit Checking

A simple transmitter control system (such as a non-processor based system) typically extends all the transmitter site meter indications to the control point, but makes no evaluations of these indications. Once a processor is added to the system, parameter limit checking can be accomplished with no additional hardware. After each parameter (or the whole group of parameters) is acquired by the A/D converter, each reading is compared with various limits. If a reading is above or below a specified limit, a specified action is taken. The software to accomplish this limit checking may reside at any point in the system (such as the transmitter site or the control point). Further, the limit checking may reside in the remote control system or may be in a computer that gathers data from the remote control system.

Remote control equipment is not legally required to have limit alarms. If, however, the station utilizes an automatic transmission system (ATS) for *attended* operation, then Section 73.1400(a)(2) requires the system to contact the operator in case of a technical problem. Early systems used analog comparator integrated circuits to detect a sample going over an alarm limit. A potentiometer adjusted the threshold voltage for the comparator. Current systems handle the problem in software (although many automatic power control systems in transmitters themselves utilize a pair of analog comparators). A typical remote control system will assign four limits to each analog channel. These limits divide the total possible range of the parameter into five bands. Limit codes may be assigned to these bands as shown Table 4.7-3 below.

A system may allow for different limit tables depending upon the state of the transmitter site. For example, different limits will apply depending on whether an AM antenna system is in day or night pattern. Different limits will apply if a different trans-

Table 4.7-3
Limit codes.

Limit	Code Description
+2	Above upper upper limit
+1	Below upper upper limit, but above upper limit
0	Inside all limits
−1	Below lower limit, but above lower lower limit
−2	Below lower lower limit

mitter is on the air. Different limits will apply if no transmitter is on the air (after station sign off, we don't care if the antenna current is below the licensed minimum!). A system could evaluate the state of the transmitter site and determine which limit table to use based on the state discovered. This may, however, cause some problems. If a transmitter goes off the air and the *Determine State* software determines that the current state is off air, it may decide to use the off air limit table (ignoring all alarms except tower lights, burglar alarm and fire alarm). For this reason, a system will often change state based only on operator commands or automatic commands (scheduled pattern changes, etc.).

Associated with each limit code and state is a *limit action*. The limit action tells the system software what to do when this limit code occurs. For example, a limit code of +1 may cause the system to trim down the transmitter power while a limit code of +2 may shut down the transmitter (if the transmitter got to this limit code, the power trim must have not been successful). A system will also typically store the previous limit code for each parameter so that transitions can be detected. A system should log the transition of a limit boundary rather than continuously printing that the power is too low. One log entry when the power went out of limits and another when the power went back in is generally sufficient. Of course, one of the limit actions will be to alert the operator of the limit transition (alarm on limit transition).

A system may highlight the out of tolerance parameter on a display, display a user specified alarm message, or some other action. In some systems the out of tolerance parameter or alarm message will flash and an audible alarm will sound until the operator acknowledges the alarm by pressing a key. Once the alarm is acknowledged, the audible alarm and flash stop, but the alarm message remains visible as long as the parameter is out of tolerance.

Time of Day Actions

If the time of day (along with day of week and date) are stored in the system as a number (generally integers), then the same limit checking routines used for analog parameters can be used for time of day automatic control. If a lower limit is set to the sunrise time, and the upper limit is set to sunset time, actions (pattern/power changes) can be generated by the limit checking software. Many systems allow for time of day programming using this or similar techniques.

Operator Interface

A wide variety of operator interface options are available. These range from the numeric display of a single metering channel along with channel select and raise/lower buttons through DTMF/voice to CRT screen displays. The single channel display is traditional (processor based systems emulating nonprocessor based systems) and quite adequate where a small number of parameters are to be monitored.

The DTMF/voice interface has proven quite popular due to the ready availability of Touch-Tone phones—a station technician already has a control point terminal at home and perhaps even in the car. With a large number of parameters to be monitored this interface can become awkward, but it is certainly better than no interface at all.

Parameter Display

Parameter display started with analog meters (either a meter for each parameter or a single meter with a function select switch). A major improvement was the introduction of digital displays in both processor and nonprocessor based systems. Voice display of parameters extends the precision and readability of digital displays to anywhere there is a telephone.

When a large number of parameters are to be displayed, it is difficult to beat a CRT screen. A typical screen would include each analog parameter along with a label and units identifying the parameter. Out of tolerance analog parameters can be identified through distinctive text (inverse video, flashing or a color such as red). Transmitter site status can similarly be displayed with a label and an indicator of the current status. An operator can view this one screen and have a good idea of the overall condition of the transmitter site. One screen may give an overview of the entire system (several transmitter sites), while additional screens give details of specific sites. Additional screens may give details of a specific piece of equipment at a specific site.

Operator Command Input

Operator command input started with the switches on the front of the transmitter. These switches may be extended to a remote location giving a switch per function. Early in the history of the remote control industry, control took the form of a channel select switch (or even a telephone dial in the very early units) with *raise* and *lower* buttons. The channel select switch acts as a function select switch, determining what the raise or lower key does. This approach continues to be used in many systems, both those with dedicated control point terminals and those that use a DTMF telephone as the control terminal.

Systems that utilize a CRT display often use an associated alphanumeric keyboard for control input. Hitting various keys sends the control functions identified on the screen or changes the screen to display additional pages of information. Use of multiple screen pages allows a limited number of keys to serve many purposes, since the function of the key varies depending upon which screenpage is visible.

Systems utilizing a CRT display may also use a *touch screen* to simulate a transmitter front panel. The screen can use boxes of varying video brightness or color to simulate back lit pushbutton controls. The operator can push the simulated transmitter button, the transmitter comes on, and the button turns green. Additional buttons provide other control functions and select other screen pages.

A control system may also utilize a computer mouse for operator control input. The operator uses the mouse to click buttons for desired commands or to select screen pages.

Automatic Logging

The FCC used to require broadcast stations to maintain an operating log, program log and maintenance log. The Commission has dropped the requirement for a program log and combined the operating and maintenance logs into the *station log*. The requirements for a station log are discussed in Section 73.1820 of the Rules. These rules include a provision for automatic logging.

Early automatic logging systems consisted of chart recorders that would plot the several transmitter parameters on a single roll of paper. This was often done by multiplexing several analog samples into a single channel recorder, then enabling and disabling the marking of the recorder as each channel stabilized. As digital circuitry was introduced to transmitter control systems, the chart recorder was replaced with a printer based on an electric typewriter. As microprocessors were introduced to transmitter control systems, standard computer printers (generally dot matrix) were put into service printing logs showing transmitter parameters.

Each of these logging systems satisfied the then existent FCC requirements for operating or station logs. At this point, required parameter logging for most stations is minimal, though a paper trail of parameter logging is still useful in showing a history of compliance with FCC requirements. Logs now serve as an indicator of drift or possible future failures in the transmission system. Having these logs as columns of numbers on a piece of paper may not be the most useful form. Further, the log printing interval may at times be too frequent (wasting paper) and at other times not frequent enough.

For these reasons, some systems now save the log to a hard drive in a form that is easy for other software to analyze. Due to the ever increasing capacity of hard drives, a relatively high log frequency can be used without generating lots of wasted paper. For example, a transmitter site might transmit a data record to all sites in the system on a user defined significant change in any parameter at the site. The record would consist of the site identification, date, time and all site parameters (typically analog, status, and other system inputs). Each site in the system captures this data to its hard drive. Each site can then build screens based on the latest data from each site in the system. The sites can also build printed logs based on the current data from multiple sites. Users can scroll back through data on each user screen, showing the history of the particular site. Finally, the log data can be exported to other software for further analysis, perhaps graphing parameters against time, emulating the chart recorders that served as the first automatic logging systems.

Programming the System

Prior to processor based remote controls, system setup and calibration consisted of adjusting the calibration pots. As features are added to processor based systems, more and more user input is required to describe how the system is to act. Keeping the system flexible while not overwhelming the user is a challenge for system designers. As with any other computer program, various methods of user data entry exist. Most systems utilize a series of menus, data tables or an adaptation of a standard computer language.

Providing flexibility in the design of user screens and printed log formats has proved challenging. Forms and menus give the user a certain amount of flexibility and ease of use. A programming language provides more flexibility and is generally less easy to use. In recent years, however, the World Wide Web has become extremely popular. The design of web pages is similar to the design of user screens in a transmitter control system. At this writing, embedded web servers are among the most popular topics at conferences dedicated to embedded system design. Simple web servers have been built into soft drink vending machines, copiers and other appliances. It is expected that transmitter control systems will also become embedded web servers, allowing the user to design screens using the standard hypertext markup language (HTML). With HTML, the user interface software becomes a standard web browser, currently available for free or at low cost. User screens can include graphics, menus, tables and other elements to make the display more understandable. These screens can be hand coded in HTML or standard HTML editors can be used to design fancy pages with little or no HTML coding by the user.

HTML pages are generally transmitted by the server in a ready to display form except for graphics. Graphics are referenced in the received page. The browser then uses these image references to request the individual graphics files. In transmitter control systems, it is not possible to have a ready built page ready to send, since the transmitter parameters vary with time. Such pages could be built on the fly by the server (a transmitter site unit), inserting current transmitter parameters into a predefined page. However, higher speed and greater user flexibility is probably available by handling transmitter parameters in a manner similar to images in standard HTML pages. The user screens can be defined as standard HTML files on the local unit, eliminating the need for transmitting the page from the server (the transmitter) to the client browser (the studio). The local page would include image tags that would request the individual transmitter parameters, similar to the way page hit counts are handled in pages on the World Wide Web. Finally, these transmitter parameters could be dropped into the page by sending individual requests to the appropriate transmitter site and waiting for the parameter to be returned over the communications link.

As an alternative, transmitter sites can send all the site parameters to all sites in the system on a significant change (as discussed in logging). The local browser responsible for building the user screens then refers to this log file for the latest data. In this manner, screens can be built quickly, as there is no wait for the data to be sent down a possibly slow and congested communications link. The studio unit is maintaining an updated cache of the information it needs, or is acting as its own *proxy server*.

Mouse clicks on screens on a web browser are generally used to bring up other pages through hypertext links. However, mouse clicks can also send commands to a program running on the server (generally referred to as a common gateway interface (CGI) script). Use of similar commands allows user interaction with the transmitter site (making adjustments, turning equipment on and off, etc.). Thus, it appears that use of a standard web browser can serve as the user interface in a transmitter control system.

Standard web browsers and the IP can be used on private networks for transmitter control. With appropriate security measures, this data can be dropped onto the Internet allowing transmitter monitoring and control from anywhere in the world at very low communications costs.

TRANSMITTER CONTROL DIVISION OF RESPONSIBILITY

Automatic transmitter control systems are often divided into several hardware boxes, each responsible for a portion of the job. Some of the control responsibility resides in the transmitter. Some resides in a remote control transmitter site unit. Some resides in a remote control control point unit. Some may reside in a control point computer.

Extended Voltmeter

The simplest remote control systems utilize an *extended voltmeter* for telemetry. In response to operator requests, the transmitter site analog multiplexer connects the voltmeter input to the various parameter samples. The operator logs the resulting readings, does limit checking and makes required adjustments. The transmitter site equipment has minimal responsibility.

Transmitter Site Polling

Automatic equipment may be placed at the control point to poll the various transmitter parameter samples (just as the operator did). This equipment may then log the readings, do limit checking and make adjustments. The intelligence of the system is at the transmitter control point and may be in the remote control unit itself, or in an external computer (through a computer interface on the remote control).

Continuous Telemetry Transmission

Another design approach calls for the transmitter site equipment to continuously scan all the parameter samples and send the resulting data to the control point. This data is then held in memory by the control point. As new data for a specific parameter is received, the old data is overwritten. The operator may then select (through a standard *channel select switch*) which parameter is to be displayed. The data is immediately pulled from local RAM. If a CRT display is used at the control point, the screen is continuously updated with the received data, giving a full screen display of all transmitter site data.

Exception Reporting

In the previously described systems, telemetry data is continuously transmitted, even if it duplicates data that was sent a few seconds earlier. Another approach is to send only the data that has changed (requiring parameter identification to be sent with each parameter), or to send the entire status of the site (all analog and status parameters) upon a significant change in any parameter. Either approach allows the introduction of noncontinuous data communications between the control point and the transmitter. The typical noncontinuous data communications link utilized is the dial-up telephone. The telephone company charges for the line during the time it is used. A dedicated circuit is assumed to be used at all times. The break even point between dial-up and dedicated voice grade circuits is typically at two to three hours per day.

By placing some of the intelligence at the transmitter site, the amount of information transmitted to the control point can be reduced considerably (through exception reporting). This allows more sites to share a dedicated circuit (due to lesser loading of the circuit), or allows the use of noncontinuous circuits, such as dial-up circuits. A transmitter site may transmit the status of the site due to any of several conditions. These include the existence of an alarm condition, a change in a parameter or a scheduled report. Systems that have dedicated circuits to the control point may immediately transmit the current state of the site based on any of these conditions. Systems that utilize dial-up circuits will typically originate a call only on an alarm condition. These systems may buffer all the parameter changes. This buffer is dumped to the control point on it filling, on an alarm condition (requiring immediate notification of the operator), and on a scheduled basis. One multisite dial-up system utilized by a television network has each transmitter site do a buffer dump each morning between 6:00 and 7:00. If all goes well, this may be the only report from the site during that day. It occurs when the telephone charges are the lowest. It also confirms that each site is operating properly as the broadcast day begins.

Data Analysis

Each week, the designated chief operator of a broadcast station is required to review the station logs and certify that the station operation was in complince with FCC requirements (FCC Rule Section 73.1870(c)(3)). In addition, the chief operator will typically look for drift in parameters in an effort to anticipate problems at the transmitter site. Statistics gathering may be done by hand (looking at manually or automatically printed logs), by the remote control (especially programmable systems) or by a separate computer. If either the remote control or an attached computer saves data to a hard drive, another program (either on the same computer or on a local area network) can access this data to build reports on the status of the system.

Data Reporting

Various devices may be used at the control point to make the transmitter site data available to the operator. These include a visual display (whether a small single channel display or a CRT display), printer, DTMF/voice interface (through a telephone) and fax machine. Visual displays show the last data received. Printers show all the data received, forming a printed log. DTMF/voice systems report the current data and alarm conditions. Systems with visual displays often include printers to generate a printed log. DTMF/voice systems typically do not generate a printed log (the Touch-Tone phone does not have a printer port). However, most stations now have facsimile machines connected to the dial up telephone network. These may be utilized by DTMF/voice systems to generate a printed log. The transmitter site equipment buffers all parameter changes, alarm conditions, operator alarm acknowledgements, automatic adjustments and operator adjustments. After each day is completed (some time after midnight), the transmitter site unit calls the fax machine and prints the log. If several transmitter sites are reporting, these calls may be scheduled to avoid conflicts. The use of the fax machine allows a station to get an automatically printed log using existing equipment.

Remote Diagnostics

The systems discussed so far have merely presented the transmitter site data to an operator. They have done minimal analysis of the data. One broadcast software company has developed transmitter analysis software. Through extensive analysis of the failure modes of transmitters, the system is able to determine what component most likely failed to cause the existing transmitter parameters. This approach is an automated *logic tree*, similar to the flow charts equipment manufacturers supply to aid in the troubleshooting of equipment.

COMMUNICATIONS CIRCUITS

Various circuits are available to connect a control point with transmitter sites. They include the following:

- Telco metallic pair
- Telco voice grade pair
- Telco dial-up voice grade
- Telco digital circuits
- STL subcarrier (voice grade or digital)
- Digital STL datastream
- FM subcarrier
- TV aural subcarrier
- TV vertical blanking interval
- AM subaudible
- Dedicated radio link (P channel)
- FM squared satellite
- SCPC satellite; Satellite VSAT
- Digital satellite datastream
- Unlicensed radio link

Telco Metallic Pair

The telephone system is designed for voice communications. Individual pairs of wires connect each telephone instrument to the local central office. Once the pair is terminated in the central office, the voice is generally digitized (8 bits per sample, 8,000 samples per second) for switching and transmission to other central offices. The links between central offices are now high speed digital lines. As such, it is very difficult to get long distance metallic pairs (dc continuity end-to-end). Within the coverage area of a single central office, however, these lines are sometimes available. They may be ordered as a *3002 local area data channel* line with dc continuity. These lines are suitable for dc metering and control. They can also be used to transmit digital telemetry and control by using short haul modems or line drivers. These condition the data for transmission over the long distance balanced lines where there is likely to be substantial common mode voltage. A low cost, short haul modem can typically transmit 9,600 bps data 2.2 miles. At 1,200 bps, this increases to 6 miles. These short haul modems normally require a four-wire circuit (one pair in each direction), though, with appropriate drivers and software, it is possible to set-up a single half duplex current loop circuit that allows communications in one direction at a time over the single pair.

Telco Voice Grade Pair

If a 3002 local area data channel goes through more than one central office, it will probably not have dc continuity. The link between central offices was multiplexed onto wire pairs, coaxial cable, microwave, or optical fiber. The telephone company can multiplex 24 voice grade circuits onto a single pair, so it is not economical for them to allocate a pair to one voice grade circuit. The voice grade circuit can, of course, carry voice DTMF, making it suitable for DTMF/voice control. It can also carry analog remote control telemetry that is based on audible tones. Finally, with the addition of modems (whether inside the remote control or outside), it can carry digital data, allowing its use on microprocessor and nonprocessor based digital control systems.

Note that the telephone company central office normally digitizes the incoming voice grade circuit at 64,000 bps. Based on this, an operator should be able to send 64 kbps through a voice grade circuit. However, due to noise distortion and delay distortion on the local loop between the customer and the central office, current modems are limited to about 56,000 bps. In one direction with 33,300 bps in the other direction over dial-up telephone lines. These modems use pulse amplitude modulation.

Other high speed digital transmission techniques are available for use over standard telephone company twisted pairs. These include integrated services digital network (ISDN) and asymmetrical digital subscriber line (ADSL). The twisted pair is not dead!

Telco Dial-up Voice Grade Circuit

In December 1984, the FCC dropped the failsafe requirement from its rules, making the use of temporary circuits legal for the control of broadcast transmitters (further, in 1995, they dropped the requirement that there be an operator present to supervise the transmitter (unattended operation) making the use of *no* communications circuit possible). These *plain old telephone* circuits are inexpensive, reliable and very widespread. They are voice grade, allowing various data encoding techniques. The most common data encoding techniques for dial-up lines are standard dial-up modems (for digital data) and DTMF/voice. The wide availability of personal computers has reduced the cost on dial-up modems, making them suitable for driving control point terminals and computers.

DTMF/voice systems utilize dual-tone, multifrequency (Touch-Tone) signals for control and synthesized (or digitally recorded) voice for telemetry and status reporting.

Use of the dial-up telephone network allows large transmitter control networks to be built, since the telephone company equipment handles all the data routing between the transmitter sites and the control points. The FCC does, however, require the broadcast transmitter operator (assuming attended operation) to have the capability to turn off the transmitter at any time. In the *Report and Order* (see 60 FR 55481) adopting the unattended operation rules, the FCC states that a dial-up line dedicated to the transmitter control system is sufficiently reliable. If, however, the line serves other purposes, an alternate means must be available for shutting down the transmitter, since emergency shutdowns cannot be scheduled for when the dial-up line is not in use. (For more information on this subject, see the *NAB Guide to Unattended Station Operation.*)

Telco Digital Circuits

The digital circuits available from telephone companies are based on the 64 kbps rate used to encode voice. Allocating a full voice grade digital circuit to broadcast transmitter telemetry severely underutilizes the circuit. However, some stations are now utilizing T1 circuits (1.544 Mbps, designed to carry 24 voice circuits) as a digital STL. With appropriate encoding, it is possible to run composite or discrete program quality stereo through a single T1 circuit. A small portion of the T1 circuit is typically allocated to be a low quality voice circuit or low speed digital circuit for transmitter control and telemetry. In each case, the 1.544 Mbps circuit is broken into several lower speed data channels suitable for carrying the digitized audio and/or other data.

Optical fiber is being utilized throughout the telephone industry. It is typically used between central offices to handle very high speed digital data (generally representing many voice circuits). At this time, most telephone subscribers do not require sufficient data bandwidth to extend the fiber the last mile to customer premises. The one possible exception is where television stations utilize digital fiber for the transmission of digitized video. As with T1, some data capacity can be stolen from the digitized video channel to provide a digital data circuit between the studio and the transmitter, suitable for transmitter control and telemetry.

STL Subcarrier

Many stations use an aural or television STL to get programming from the studio to the transmitter. There is usually bandwidth available for the addition of a voice grade or digital data channel. In addition, television stations may insert control data in the vertical blanking interval of the video going to the transmitter site. This data may be replaced with telemetry data prior to being fed to the broadcast transmitter. The most common technique of sending control data over an analog aural STL is to use an FM subcarrier to establish a voice grade circuit. Standard modems (inside or outside the remote control) then drive this new voice grade circuit. Other modulation techniques are available, including FSK of the subcarrier, and quadrature amplitude modulating the subcarrier with digital data. The following frequencies are typically used:

- 26 kHz: Monaural or discrete stereo aural STLs
- 110 kHz: Composite stereo aural STLs
- 135 kHz: Composite stereo aural STLs
- 152 kHz: Composite stereo aural STLs
- 185 kHz: Composite stereo aural STLs
- 7.5 MHz: Video STLs
- 8.3 MHz: Video STLs

Digital STL Datastream

Digital STLs (whether for radio or television) provide a high speed digital link from the studio to the transmitter. Some of this bandwidth is used to transmit program material to the transmitter site after the programming has been encoded to a suitable digital form. Additional data bandwidth is generally available for a low speed digital channel or voice grade audio channel to send control information to the transmitter site. In transmitter to studio link (TSL) and intercity relay (ICR) applications, these links can also carry telemetry information.

FM Subcarrier

Transmitter site telemetry may be returned to the control point (and anywhere else there is a subcarrier receiver) over a subcarrier on the FM broadcast channel. This is generally a voice grade FM subcarrier driven by a modem in the remote control. Other modulation techniques are available, including FSK and quadrature amplitude modulation. In addition, the subcarrier is sometimes further multiplexed by using the 20–30 Hz range to return the metering (either analog or digital) and the range above 50 Hz for other programming. Typical FM broadcast subcarrier frequencies are (assuming stereo operation) 67 kHz and 92 kHz.

TV Aural Subcarrier

In 1971, the FCC authorized TV stations to use an aural subcarrier with an instantaneous frequency between 20–50 kHz. Much of this spectrum is now taken by BTSC stereo. BTSC does, however, set aside a subcarrier frequency for station use (the PRO channel). This is at 6.5 times the horizontal scan rate. It is frequency modulated with a voice grade signal, or may be frequency shift keyed with a digital signal.

Vertical Blanking Interval

FCC Rule Section 73.682(a)(21) allows lines 17 through 20 (except line 19, which is reserved for the standard ghost canceling signal) of both fields for test, cue, control and identification signals. On an STL, these may be used to send transmitter control information. The data could be replaced with telemetry data at the transmitter site.

AM Subaudible

In 1969, the FCC authorized AM stations to use frequencies below 30 Hz on the AM carrier for telemetry. The modulation level of the telemetry signal was limited to 6%. Analog metering systems typically use 20 Hz to 30 Hz (20 Hz representing zero, 30 Hz representing full-scale). Subaudible signaling may also be used to transmit digital data (although at a low rate, on the order of tens of bps). The subaudible digital signaling may use any of various modulation techniques. FSK and phase shift keying have been used. The low speed capability of the data channel may be a limiting factor, but over a full day, a lot of data can be sent. If the telemetry system sends only the changes in parameters while the control point terminal stores the most recently received data (perhaps using a standard CRT terminal), a reasonable telemetry system is possible.

The current FCC rule Section 73.127 governs this multiplex use of AM carriers. It allows the use of either amplitude or phase modulation (the original rule required the use of AM). The rules no longer set a maximum injection level or frequency, though the multiplex signal must not be audible or cause radiation outside the channel. AM multiplex signals are also finding other applications. One such application is electric utility load management. The subaudible band is also used by the AM stereo system for a pilot signal to trigger the operation of stereo demodulators. These other applications may limit the use of this range of frequencies for transmitter telemetry.

Digital Television

As digital television is introduced, various other circuits become available for the transmission of control and telemetry signals. Both the digital STL and the digital television transmitter provide a high speed digital circuit that may be divided into various lower speed channels. Further, since both video and telemetry/control data tend to be *bursty* (the peak bandwidth is higher than the average bandwidth), higher transmission rates are available for the various services through the use of statistical multiplexing or use of data packet transmission instead of allocating fixed data rate channels to each application.

Both the STL and the broadcast transmitter datastreams may make some data capacity available for telemetry and control.

Broadcast Subcarrier Limitations

Broadcast subcarrier (whether AM subaudible, FM or TV aural subcarrier, TV VBI, or digital TV subchannel) telemetry transmission is low cost, but has one serious limitation. When there are problems with the broadcast transmitter, no telemetry is available, making remote diagnostics and troubleshooting difficult. For this reason, many prefer to use a telemetry link that will continue to operate on failure of the broadcast transmitter.

Dedicated Radio Link (P Channel)

The FCC has set aside eight frequencies in the 450–455 MHz area for operational communications, including telemetry and control (see FCC Rules Sections 74.402(e)(9), 74.462(b), and 74.482(d)). It is most common to treat these radio links as a voice grade circuit, then use standard voice grade modem techniques to send data. It is also possible to directly digitally modulate the UHF carrier using any of the standard techniques (FSK, etc.). Many stations leave the transmitter up continuously, making the frequency unavailable to other stations in the area. Simple circuitry can be added between a standard remote control and the TRL transmitter to only key the transmitter when readings are being taken. Such systems could also use the dial-up technique of holding all the data changes in a buffer until the buffer fills, an alarm condition occurs or a scheduled report time.

Another technique available for frequency sharing is *packet radio*. Here, a terminal node controller (TNC), packet modem or packet assembler/disassembler (PAD) coordinates the use of the RF frequency, adds addressing to the packets (setting up *virtual circuits* between sites in the system) and adds error correction. The RF packet communications system provides a transparent serial digital circuit between two points over a shared RF channel. Some remote control equipment includes packet hardware and software in its design. Other systems may be adapted to packet radio by adding an external packet modem. Finally, note that Section 74.482(d) outlines the specifications for automatically identifying the P channel radio link. Some remote control systems include the ID system. Some P channel transmitters also include an ID system.

FM Squared Satellite

Some satellite networks providing transmitter operating services rely on dial-up telephone lines for routine control and telemetry of transmitters. To provide an independent method of shutting down transmitters on the failure of the dial-up system, the networks may use subaudible tones on the program channel or interrupt the program audio. This is typically used on *FM squared* satellite channels where the program audio is carried as an FM subcarrier above video (which is also frequency modulated on the satellite, although some systems delete the video, just running an FM carrier with several FM subcarriers). Since subaudible data channels can only handle low data rates, some FM squared systems run a separate data channel (on another video subcarrier). This data channel normally handles addressed packet data. It includes program cues, individual station transmitter commands and text messages to subscriber stations (electronic mail). Loss of the data channel may be used as a failsafe way of shutting down the transmitters.

SCPC Satellite

Single channel per carrier (SCPC) satellite systems run several FM audio carriers in a single video transponder, as opposed to carrying the audio as subcarriers on a single video carrier. The bandwidth of the SCPC carriers normally allows the addition of a subcarrier immediately above the program audio for each channel. Through the use of data packets and addressing, this data channel may be used for program cues, individual station transmitter commands and text messages to subscribers. Again, routine transmitter control and telemetry is accomplished over dial-up telephone lines, since each station has a receive-only earth station.

Digital Satellite

Many program networks have changed to digital distribution of their programming. A portion of the high speed datastream utilized for audio distribution can also be utilized for program control (local ad insertion, etc.) and transmitter control. This one-way link can be supplemented with a dial-up telephone as a return link for telemetry.

VSAT Satellite

Very small aperture terminal (VSAT) satellite systems are widely used for point-to-point and point-to-multipoint data distribution. These differ from the previously discussed satellite systems in that a transmit/receive terminal is put at each station. Through data packet switching, the VSAT system sets up a large number of virtual circuits through a single satellite transponder, on a single frequency. The techniques are similar to those used in packet radio. The single RF channel is divided among several users through time division multiplexing. However, instead or allocating a fixed time slot to each user, a time slot large enough to handle the data a site needs to transmit at that particular time is allocated (as in statistical multiplexing). The VSAT system appears to the user as a very long RS-232 cable. Data put in one end comes out the other. The complexity in the middle is handled by someone else. This very long RS-232 cable allows the control terminal or the transmitter control system to be located almost anywhere. To decrease the amount of time utilized on the satellite circuit, the remote control (or an interface computer) uses exception reporting, reporting only changes in system parameters (and often updating the data every half hour). Sending the same data over and over again (as would result from a stable transmitter) wastes time on the satellite that would otherwise be available to other users. The two-way satellite link allows the complete control and telemetry of the broadcast transmitter to be handled by the satellite circuit. The previously described receive-only satellite systems relied on dialup telephone circuits to get telemetry back to the control point.

Unlicensed Radio Link

Part 15 of the FCC Rules sets aside various bands of frequencies that can be used by low power communications devices without any licensing. Users of these frequencies are not protected from interference. Some stations have utilized Part 15 spread spectrum digital radio links to provide control and telemetry communications between sites. These systems can work well in uncongested areas. In other areas, the shared use of the frequencies limits their usefulness.

MANY OPTIONS

Recent revision of the Communications Act of 1934 and the subsequent elimination of the FCC requirement that an operator be in charge of a broadcast transmitter is a radical change from the traditional approach to the supervision of broadcast transmitters. Further, relaxation of the ownership limits can result in a single licensee owning several stations in one market and hundreds of stations in the country. These licensees may choose to run the transmitters unattended relying upon transmitter stability, automate the supervision of the transmitter or have a person keep a close watch on transmitters throughout the country. The degree of supervision is now up to the station licensee, *as long as the transmitters continue to operate within licensed parameters.* Stations may use the techniques discussed in this chapter (as well as other techniques) to maintain the degree of supervision desired.

4.8
AM BROADCAST ANTENNA SYSTEMS

CARL E. SMITH

INTRODUCTION

Standard broadcast (AM) antenna systems can reach a high degree of sophistication, much of which is based on advanced mathematics. The purpose of this chapter is to provide the station engineer with an understanding of some of the basic concepts of antenna design and an appreciation for the complexities of this specialty. An extensive bibliography is included for those who wish to pursue further study.

The chief purpose of a broadcasting antenna system is to radiate efficiently the power supplied to it by the transmitter. A simple antenna can do this job quite well. This is often a single vertical tower that radiates its signal equally in all directions along the ground in a so-called nondirectional or omnidirectional pattern. A second purpose of an AM antenna system is often to concentrate the power in desired directions to cover populated areas and to suppress it in other directions to protect the coverage of other stations sharing the same or closely-adjacent channels. This directionality may require a very complicated antenna system with several towers if the requirements are stringent.

The antenna is the last point in the system under the control of the broadcaster. The signals radiated from the antenna are propagated through space to each receiving antenna. The factors affecting the strength of the received signal include the strength of the signal radiated by the broadcasting station in a particular direction, the distance to the receiving site, losses incurred by the less-than-perfect conductivity of the ground along the propagation path, terrain obstructions, (large hills cast shadows even at AM frequencies) and, in the case of skywave transmission, the ionospheric conditions that determine how much of the radiated signal will be reflected back to each distant receiving location. Signal strength in a particular direction can also be affected by the presence of structures such as buildings or towers near the radiation system.

The polarization of the transmitted waves is also a factor; for standard broadcast stations vertical polarization is used because of its superior groundwave propagation and the simplicity of antenna design. The FCC has established maximum transmitter power limits for each of the four classes of AM channels—A, B, C and D so the only variables available to the design engineer attempting to maximize the coverage of a radio station involve the antenna location, the pattern design, and a limited choice of power levels. These factors go hand in hand when designing a directional antenna system. Severe constraints are usually imposed on transmitter site selection because of aeronautical, zoning, environmental, and coverage requirements. The constraints encountered in the pattern design relate to the size and shape of the transmitter site, the extent to which the necessary signal suppression can be achieved at the desired transmitter power level and the cost of design, construction, adjustment, and maintenance of multi-tower systems. The pattern design can also seriously affect the stability, efficiency and bandwidth of the completed system. These factors will be discussed later.

RADIATION VERSUS FIELD STRENGTH

Two independent factors determine the signal strength at any given point within a station's service area. First is the strength of the signal radiated in that direction; second is the path attenuation between the transmitting and receiving antennas. Attenuation is determined by both distance and ground conductivity. It is customary to express the radiation in units of millivolts-per-meter at 1 km unattenuated. This is the field that would exist at 1 km over perfectly conducting earth. In this case the field strength would be inversely proportional to the distance from the transmitting antenna; hence, the radiation is also described as the "inverse distance field." The unattenuated radiation cannot be measured directly but can be inferred with great accuracy if sufficient field strength measurements are made to determine the ground conductivity. Field strength measurements are always dependent on radiation, distance, and ground conductivity.

THE SINGLE TOWER NONDIRECTIONAL ANTENNA

The majority of single-tower antennas are neither top-loaded nor sectionalized and most of them are insulated from ground. For such simple towers, the current is a maximum 90 electrical degrees down from the top (or at the base if the tower is shorter than 90° in height). A typical guyed tower that is 90° high physically is about 95° high electrically, because the velocity of propagation is less in the tower than in air and is a function of the tower cross-section, slowing down as the cross-section is increased. The approximate shape

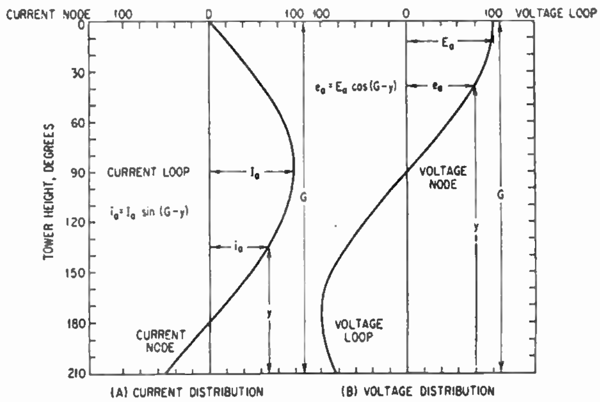


Figure 4.8-1. Theoretical current and voltage distribution on a vertical radiator.

of the current distribution on a thin tower of uniform cross-section is given by

$$i_a = I_a \sin(G - y)$$

Where: i_a is current in amperes at height y
I_a is the maximum current in amperes
G is the tower height in degrees
y is the height in degrees of the current element i_a

As an example, the general shape of the current and voltage distribution on a thin tower 210 electrical degrees high is shown in Figure 4.8-1. For shorter towers, the distribution would approximate that shown, but with the lower portions cut off; there always being a current node and a voltage maximum at the top of any such tower that does not employ top loading. It is important to visualize the shape of the voltage distribution along the tower because of the need of good insulators at the high-voltage points. Otherwise corona or arc-overs may result to disrupt broadcasting service.

The tower current and voltage are not zero at the nodes shown along the tower. Rather, they reach minimum values and shift rapidly approximately 180° in phase in traversing the node region. When towers considerably taller than 180° in height are considered, the current near the base is in the opposite direction from that in the upper portion of the tower. Under these conditions, when viewed in the horizontal plane, the radiation from the lowest part of the tower is canceling a portion of the radiation from the part above the current minimum. Any increase in tower height above the optimum would actually reduce horizontal plane radiation.

VERTICAL RADIATION CHARACTERISTICS

Maximum groundwave radiation occurs for a tower 225 electrical degrees high (5/8 wavelength). The variations in tower current distribution with increasing tower height defines the shape of the radiation characteristic in the vertical plane. Figure 4.8-2 shows the size and shape of the vertical plane radiation patterns for a single tower of various heights atop a perfect ground system, fed with 1 kW of power.

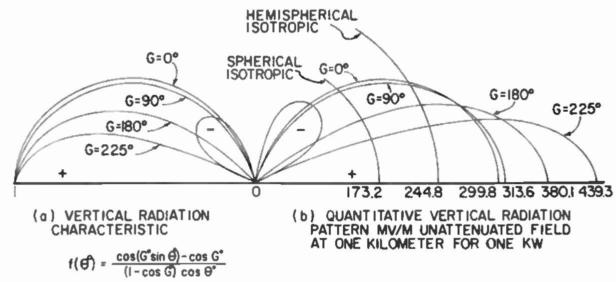


Figure 4.8-2. Radiation characteristics in vertical plane.

Insulated Tower Base Impedance

The base impedance of a single nondirectional tower is determined principally by its electrical height, its cross-section, the extent of the ground system, and the elevation of the feed point above ground. For typical guyed towers of uniform cross-section, which are base insulated and fed 4 or 5 ft above ground level, the resistive and reactive components of the base impedance approximate the values shown in Figure 4.8-3. The base impedance of self-supporting towers departs radically from the values shown, not only because of their large and tapering cross-section, but also because of the capacitance of the base insulators necessary to support each leg of such towers.

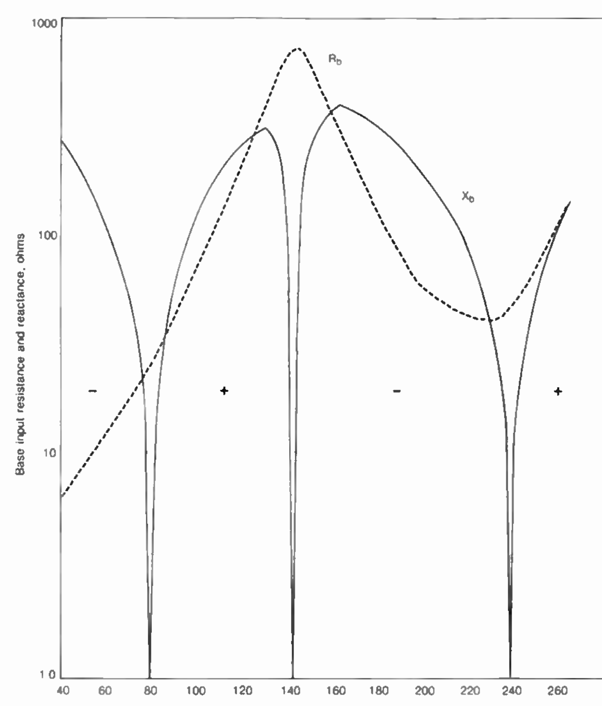


Figure 4.8-3. typical base input resistance and reactance of a uniform cross-section base insulated guyed tower.

Electrically short towers are inefficient radiators, not only because of the shape of their vertical radiation characteristics as shown in Figure 4.8-2, but also because of proportionately higher ground losses. For example, a tower 48° high with a base resistance of only 9 Ω will have approximately 10% of the available power wasted in the ground system resistance losses (typically 1 Ω).

Grounded Towers, Shunt Fed and Folded Monopoles

Occasionally towers without insulated bases must be utilized as AM radiators. Such structures include FM or TV towers, water tanks and ornamental flag poles. Although the impedance at the base of such a tower is necessarily essentially zero, the impedance rises with increasing height of the feed point. It is a simple matter to determine experimentally the height at which a shunt fed tower must be driven to provide a desirable input impedance. A common technique is a *slant-wire*[1] feed in which a wire is attached to the tower at a selected height above ground and brought down to near ground level at an angle approximating 45°, to serve as the antenna input terminal. A slant-wire feed distorts the otherwise omnidirectional pattern of a single tower and tends to suppress radiation over the sector on the side where the slant-wire is attached. This effect can be avoided if, instead of the slant-wire, the feed conductors are insulated at the base, brought up outside of the tower and bonded to the tower, for example, 90° above ground to form a folded monopole. The conductors, in this concentric arrangement, in effect form the outer conductor of a coaxial transmission line with a short to the tower at the 90° point and an open at the base insulators. This quarter-wave open circuit transmission line in effect puts an insulator at the tower base. The current up on the outer conductors and down on the tower essentially cancel so far as radiation is concerned. The tower with this insulated skirt performs like a base insulated tower. The concentric arrangement of conductors, usually six, are tied together above the conductor base insulators and fed like a base insulated tower. There is a small amount of power loss in the 90° concentric transmission line shorted at the top and used to produce the open circuit at the bottom. The radiation current is up on the outer conductors, to where they are connected to the tower at the 90° point, and then on up to the top of the tower where the current is zero as on a base fed insulated tower.

Folded Conical Monopole

The folded conical monopole shown in Figure 4.8-4 broadbands the input impedance. For a 90° tower the vertical radiation characteristic is that of a 75° tower but with an increased value of input resistance and reactance of about 50%. These results were obtained by applying the method of moments to compute the base impedance, current distribution and vertical pattern.

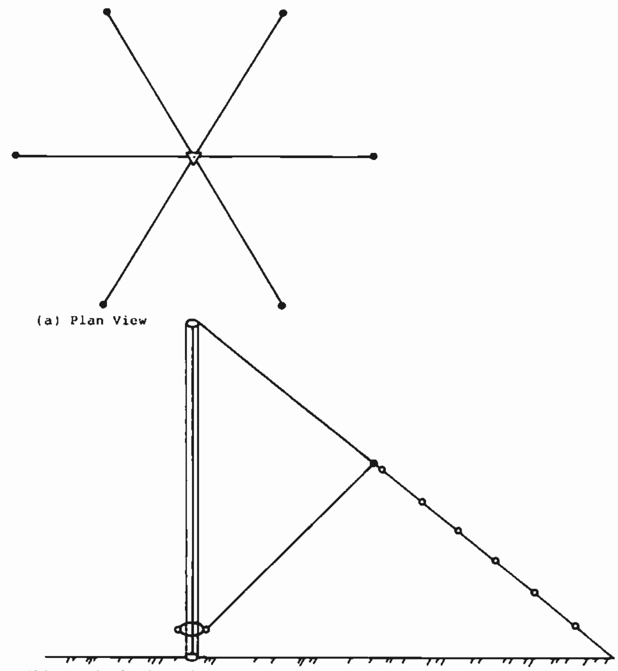


Figure 4.8-4. Sketch of folded conical monopole antenna.

Top Loading

The performance of an electrically short tower (significantly less than 90°) can be improved, both as to radiation efficiency and bandwidth by means of top loading. This consists of increasing the capacitance to ground from the top of the tower. This loading can take the form of either a flat, more-or-less circular horizontal disk attached to the top of the tower (called a top hat) or as sections of guy wires bonded to the top of the tower and extending down a useful distance before encountering the first of the guy wire insulators. Many variations are possible. Some installations use 3, 6, or even 12 nonstructural guys for top loading that are very effective. By interconnecting the lower ends of the top loading cables, the capacitive loading is increased some, but this method or spider web connections between the top loading cables increases the construction and maintenance problems. These problems can be eliminated by just increasing the top loading cables a small amount to give the same increase in capacity effect. Top loading is electrically less desirable than increased tower height, but is useful where towers must be electrically short due to either extremely low carrier frequencies or to aeronautical limitations. Top loading increases the base resistance and lowers the capacitive base reactance thus reducing the Q and improving the bandwidth on towers less than 90° high. When the tower height is of the order of 130°, top loading can be used to increase the tower's

[1] Slant wire feeds are no longer accepted by the FCC for critical hours and night operation.

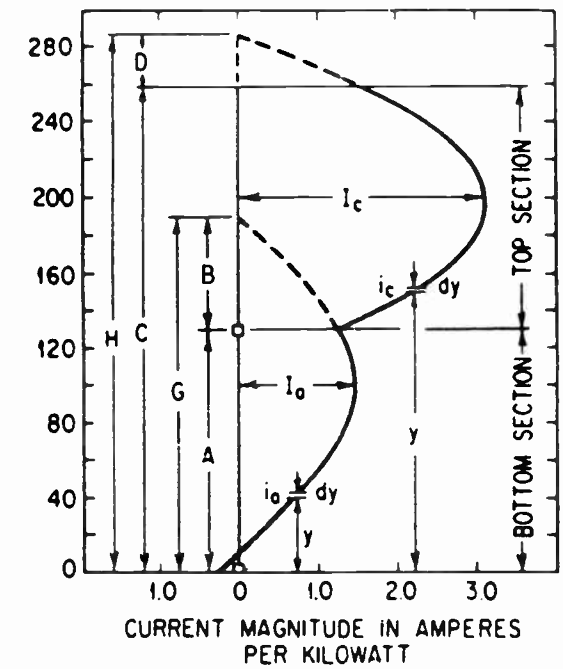


Figure 4.8-5. Theoretical current distribution on top-loaded sectionalized tower.

electrical height to give maximum groundwave radiation and minimum skywave radiation.

Sectionalized Towers

A utopian vertical radiator would have a constant current throughout its height, but in real life the current must ultimately reduce to zero at the tower top or at the end of the top loading cables. The current can be made to diminish less rapidly by inserting an inductance in series with the tower at a point part way up its height. This is the same technique as the familiar "loading coil" near the center of the vertical whips often used for mobile radio systems.

Top Loaded Sectionalized Tower

For a simple vertical radiator the radiation characteristic can be improved by increasing the tower height up to 225° for maximum ground wave or by top loading. This in effect raises the position of the current loop with respect to the ground. This principle can also be applied to the top section of a sectionalized tower.

The purpose of top loading a sectionalized tower is to provide a means of further controlling the current distribution on the lower section only. Considering efficiency and stability, it is usually possible to achieve a more favorable radiation characteristic of the whole tower by employing top loading and sectionalization (see Figure 4.8-5). In the case of tall towers used to support FM or TV antennas, it may not be practical to employ top loading.

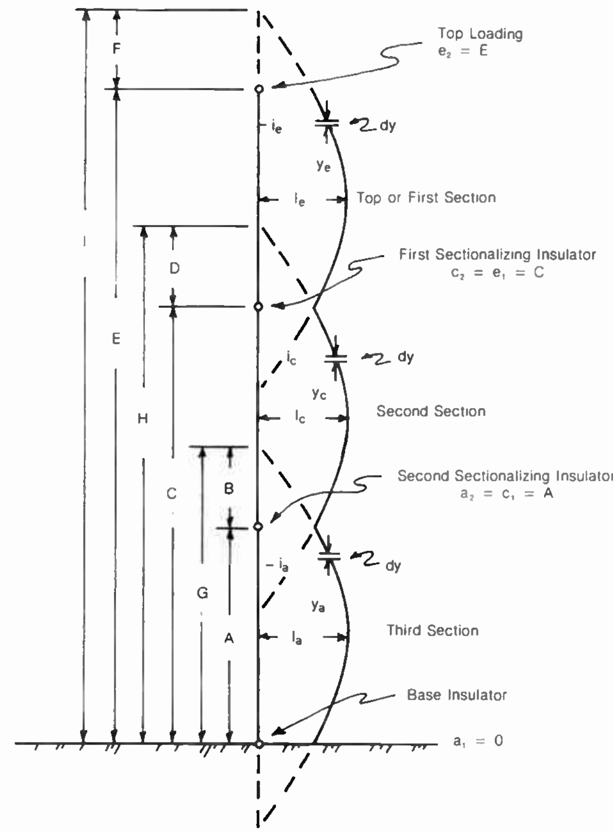


Figure 4.8-6. Theoretical current distribution on three section top loaded tower.

Depending on the height, of tower in wavelengths, the tower can be sectionalized at one or more points to accomplish the highest efficiency consistent with good operating stability (see Figure 4.8-6).

Short Low-Loss Antenna with Insulated Counterpoise

Short low-loss antennas are useful for standby use and regular use where height is limited. By using optimum top loading and a tuned counterpoise, the field strength is maximized by adjusting the counterpoise inductor in Figure 4.8-7 to minimize the top hat field through the counterpoise to the lossy ground.

Ground Systems

The current on a tower does not simply "disappear," rather it returns to earth through the capacitance between the earth and each incremental element of the tower or the top loading. For towers not exceeding 90° in height, the tower current is greatest at the base. For such towers the radial ground current is greatest near the tower and decreases with increasing distance from the tower. For single towers the ground currents are radial from the tower base. The ground losses are greatly reduced if the tower has a radial copper ground system, so the ground current will be in the low-loss

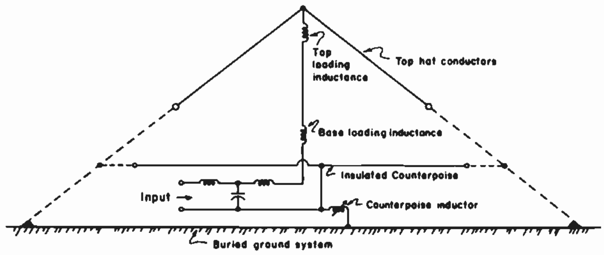


Figure 4.8-7. Short, low-loss antenna with counterpoise.

copper ground system rather than in the earth which has a much higher resistance. A solid copper sheet of infinite radius would be the ultimate ground system, but experiments and experience have defined the dimensions of an adequate ground system. A system of 120 radial ground wires, each 90° long (140° is considered optimum), and equally spaced out from the tower base, constitutes a "standard" ground system. This is often augmented with an additional 120 interspersed radials 50 ft long, or an expanded copper mesh ground screen 25 to 50 ft square centered at the tower. A superior ground screen material is the copperweld, mesh ground mat often utilized by power companies for lightning protection under electrical substations.

Where the antenna site is too small to accommodate all the ground radials at full length, a compromise often used, if easement can not be obtained beyond the property line, is to increase the number of radials by placing them 1 or 2° apart rather than the standard 3° separation.

There is no magic in a "standard" ground system for nondirectional towers; it simply represents a reasonable balance between cost and radiation efficiency.

The antenna system loss including the tower and ground system is normally assumed to be 1 Ω and is added to the tower base resistance.

Most ground systems under directional antenna arrays consist of the usual 120 radials per tower truncated and bonded to traverse copper straps where the radials from the several towers would otherwise intersect. Stability considerations may dictate larger than standards ground systems under critical directional antenna arrays; changes in soil conditions beyond the ground system can result in small changes in tower base impedance.

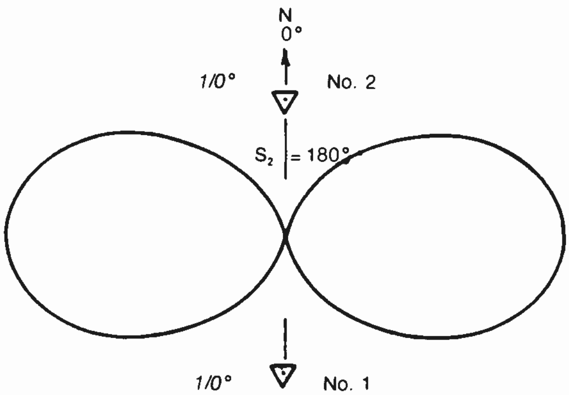
Ground system losses are minimized if the radial wires are placed above ground, thus the E-field voltage from the tower and top loading cables terminate on these radial conductors so the H-field current can return to the tower base without penetrating the lossy earth. Ground radials are usually buried 6 to 8 in. for mechanical protection. Burial up to 24 in. is feasible where necessary to permit deep plowing for agricultural crops. However, the ground system should be very near the earth surface in the immediate vicinity of the tower. The earth losses are greater for the buried ground system. Changes in weather conditions change the dielectric constant and conductivity of any unshielded earth to the detriment of base current stability.

TWO-TOWER DIRECTIONAL ANTENNA

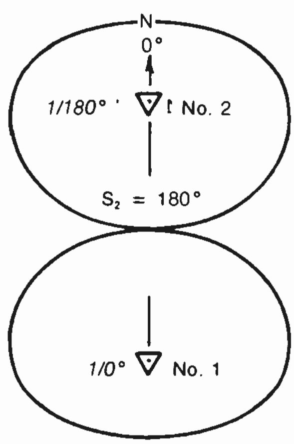
When a nondirectional antenna, with a given power, does not radiate enough field strength to serve the community of interest and/or fails to protect other radio stations, then it is logical to resort to a directional antenna system to achieve these objectives. FCC Rules spell out the protection requirements to be provided to the various classes of stations, both daytime and nighttime on the same and adjacent channels. These limits, which must be met in the directional antenna design, tend to define the shape and size of the most desirable antenna pattern. Since the distances and directions to the other stations requiring protection are rarely the same, most directional antenna patterns are tailored to meet the specific requirements. A directional antenna functions by carefully controlling the amplitude and phase of the radio frequency currents fed to each tower. The resulting field in any direction is the vector sum of the individual tower radiation components. To visualize the resulting pattern in the horizontal plane, one must consider the individual tower radiation components when viewed from distant points in different directions. The relative amplitudes from the individual towers remain unchanged, but the relative phases shift with azimuth because the signal from the closest tower arrives first. In a directional antenna system, one tower is usually defined as the reference tower, and the amplitude and phase of each other tower is measured relative to this reference. The reference tower usually has the greatest current, thus the ratio of the current in each other tower relative to the reference tower current is a fractional number often expressed as a percent of the reference tower current. The relative amplitude and phase of the tower currents is measured by means of an antenna monitor.

The phase of the field, radiated by each tower relative to the reference tower, has two components when viewed from any distant point of observation. The relative electrical magnitude of the current fed to the tower is one component, and is adjustable. The second component is the phase which appears to lead or lag the reference tower by virtue of being more distant or closer than the reference tower to the point of observation. This is termed the space phase component and varies continuously for each tower in a sinusoidal manner as the observation point is moved in azimuth along a distant circle around the array.

Figure 4.8-8 shows three simple directional antennas and their resulting patterns which are easy to visualize. Figure 4.8-8(a) shows two towers arranged along a north-south line separated by 180° and fed with equal currents in phase. When viewed from the east or west, the fields from the two towers are in phase and the maximum field strength results. When viewed from the north or south, the field from the more distant tower is delayed by the 180° of additional distance,



(a) Broadside Array



(b) End Fire Array

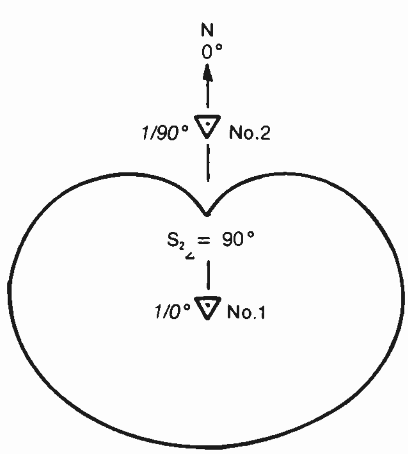


Figure 4.8-8. Three simple directional antenna patterns.

thus canceling the field of the closer tower so as to result in a minimum or null. The deepest minimum or null occurs only when the fields are exactly equal in amplitude and opposite in phase.

Figure 4.8-8(a) is termed a broadside array because the maximum radiation is broadside to a line through the towers. Figure 4.8-8(b) shows a similar arrangement, but with the phase of the current in the north tower shifted by 180°. Then the fields from the two towers cancel each other when viewed from the east or west, but would produce maximum radiation from north or south. This would be termed an end-fire array, because maximum radiation coincides with a line through the ends of the array. Figure 4.8-8(c) alters the spacing to 90° and phasing to 90° so as to produce a cardioid pattern. Other combinations of tower spacing and phasing can produce a great variety of pattern shapes. See Appendix B.

Multiplication of Two Tower Patterns

Perhaps the most widely used method of controlling pattern shape involves the multiplication of two-tower patterns. This illustrated in Figure 4.8-9. When a two-tower pattern such as pattern Number 1 with nulls at $\pm 0_{n1}$ is multiplied by pattern Number 2 with nulls at $\pm 0_{n2}$ the result is pattern Number 3 in a three-tower array. The directions of all of the two-tower array nulls are maintained in the three-tower array. This is a very powerful design technique for protecting other stations and still serving a desired service area. In this special case the spacings S_2 and S_3 are equal, resulting in an in-line array with fields of towers Number 2 and Number 3 being added in the center tower, and the end tower of the three-tower array is the multiplication of these fields as shown in pattern Number 3 of Figure 4.8-9.

In the event that the protection directions are not symmetrically located, the two-tower arrays can be placed on different azimuth angles to produce a four-tower parallelogram array (as shown in Figure 4.8-10). The nulls of the Number 1 pattern are maintained and the nulls of Number 2 pattern are maintained in the four-tower parallelogram array. Furthermore, the spacing from Number 1 tower to Number 2 and Number 3 towers does not have to be the same. By this approach of using one or more parallelograms, a wide variety of asymmetrical patterns are possible with relative simplicity of pattern calculations. However, modern computer techniques can optimize individual tower locations, currents and phases so as to produce an efficient pattern, frequently using fewer towers than required with the parallelogram approach.

Systematization of Patterns

The pattern possibilities resulting from variations in spacing and phasing have been systematized and a sample of two-tower patterns is shown in Appendix B. (See Reference 1 for three-tower patterns.)

Radiation Pattern Size

The pattern size is usually determined by integrating the energy flow outward through an imaginary hemispherical surface surrounding the directional antenna array. This method does not give information regarding the distribution of power radiated from the various towers of the directional antenna array, however, it is very useful for making comparisons of pattern size.

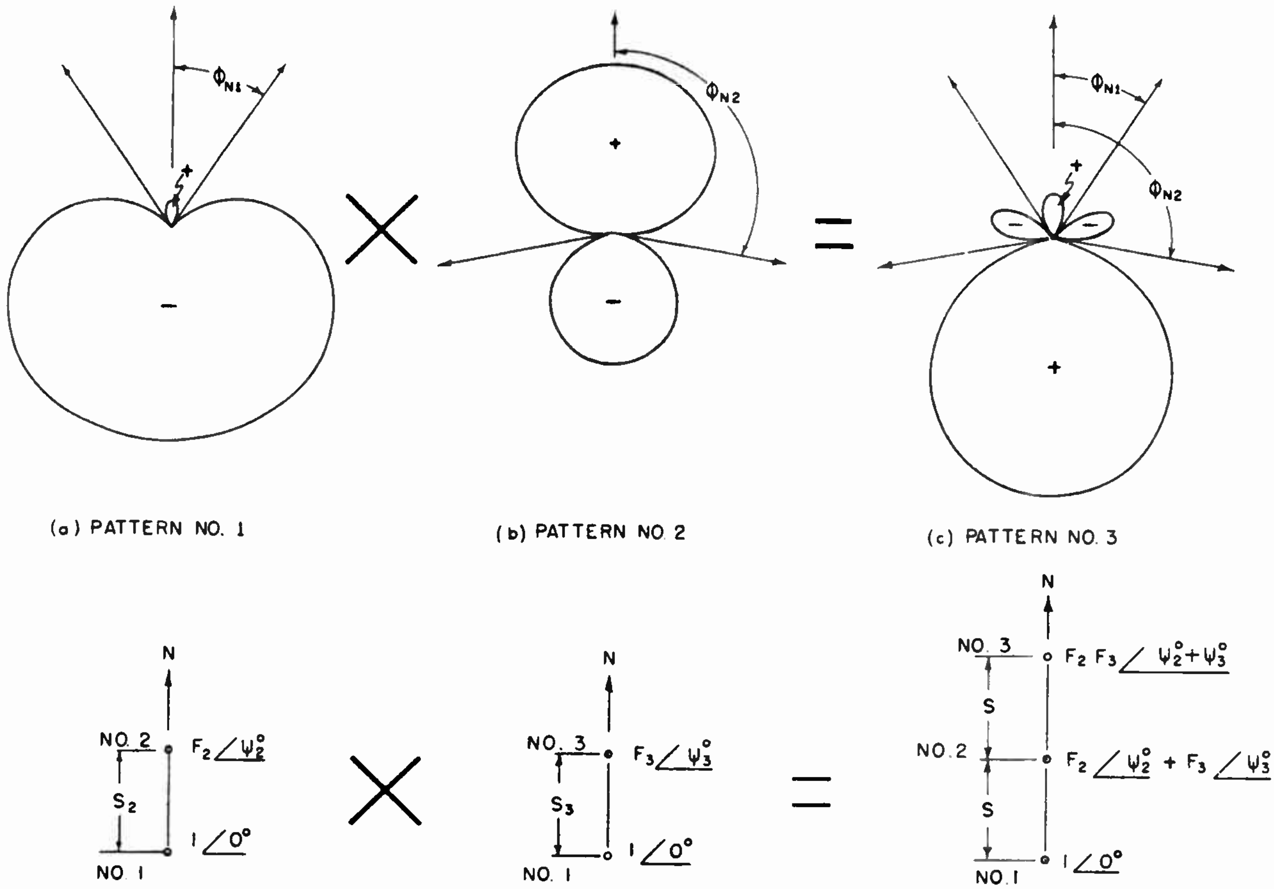


Figure 4.8-9. Multiplications of patterns to produce a three tower in line array.

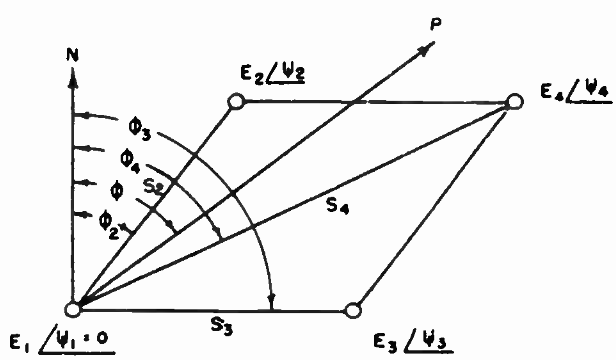


Figure 4.8-10. Multiplication of two patterns to produce a four tower parallelogram array.

This computation method is available in digital computer programs and is used by the FCC.

There are other methods of determining pattern size, such as the *mutual resistance method* which employs Bessel functions, and the *driving point impedance method* which uses mesh circuit equations with self and mutual impedance information.

The "method of moments" is now available in large computer programs to determine current distribution on towers and top loading cables, base driving point impedances and the vertical pattern of directional antenna arrays.

Driving Point Impedance

The input impedance of each tower in an array (*driving point impedance*) is not that of the nondirectional tower. The driving point impedance contains the self impedance plus the mutual impedance multiplied by the current ratios that exist in the array as driven to produce the desired pattern. The driving point impedance will modify the self impedance, depending on the array parameters, and can even make the base resistance negative so that the tower draws power, from the other towers and dissipates the power into a load resistor or delivers it back to the phasing system. Because the driving point impedance is affected by the currents in the other towers, it can only be measured by an operating bridge inserted in the tower feed point, while the other towers are operating with their correct current magnitude and phase.

Base Currents Versus Radiated Fields

In a directional array, the tower base current ratios will usually depart substantially from the calculated radiated field ratios when the pattern is correctly adjusted. This is caused by the mutual coupling between

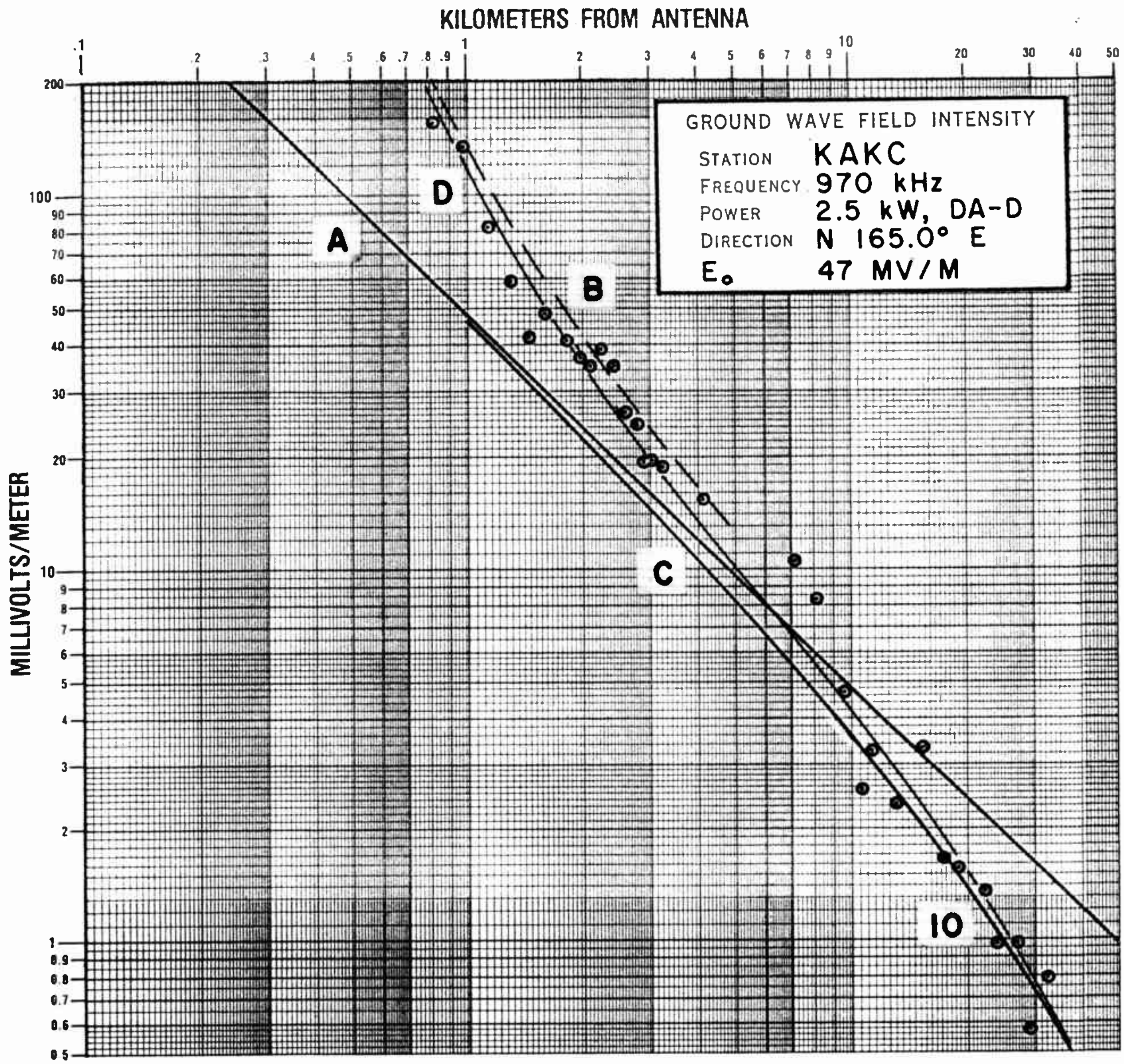


Figure 4.8-11. Near-field effects.

towers which distorts the sinusoidal current distribution otherwise assumed for each tower. Thus, the correct pattern is initially proved by means of a series of field strength measurements in significant radial directions from the station rather than by assuming that measurement of tower currents and phases can establish the correct pattern.

Near-Field Versus Far-Field Conditions

Theoretically a directional antenna pattern is not fully formed except at an infinite distance, where the separate towers can be considered as point sources. As a practical matter, near-field effects can persist as far as 32 km (20 miles) from an antenna before far-field conditions prevail. This is especially true in the deep minimums of wide-spaced arrays; however, misleading measurement results can often occur under apparently innocent circumstances. Near-field calculations involve consideration of the actual inverse distance attenuation and the actual phase delay from each antenna element to a series of observation points along a radial.

Figure 4.8-11 shows the results of such calculations on a minimum radial and the resulting analysis of field strength measurements. Line A is the inverse distance line for the theoretical unattenuated radiation at 1 km. Line B is the result of the near-field calculations assuming only inverse distance attenuation, that is, no soil losses. It converges with the inverse distance line with increasing distance. Line C represents a soil conductivity of 10 mmhos/m as drawn in the conventional manner from analysis of nondirectional measurements on the radial. Line D is a composite of lines B and C. It includes the near-field calculations and is attenuated with distance in accordance with the soil conductivity previously established. This composite line converges with the near-field calculations at short distances where soil attenuation is negligible and converges with the soil conductivity line at great distances where near-field effects disappear. Since curve D accounts for both near-field effects and soil losses, it is the proper curve against which the directional field strength measurement data should be fitted. Note the good fit to the measurement data, both close to the array and at distant points even though the first 19 measurement points fall considerably above the inverse distance line A.

Pattern Size Versus Pattern Shape

The shape of a directional antenna pattern is determined by the adjustment of the phase and ratio parameters, whereas the pattern size is a measure of the power radiated and is affected by the transmitter power output and the losses within the phasing system and the ground system. Since pattern size and shape are essentially independent, it is most expeditious to adjust an array to get the correct shape before expending much concern on the size.

Field strength measurements on a previously-licensed directional antenna may appear to indicate a change in pattern shape or size when the change was in fact due to changes in soil conductivity. Such changes affect distant measurements more than close-in measurements. In some areas of the United States, the conductivity is typically higher during winter and spring months when the soil is more moist than in summer and fall months.

Seasonal conductivity variations are not observable in some portions of the country, yet are extreme in other areas. One well-documented case showed a seasonal doubling of signal strength at 32 km (20 miles) in the main lobe of a correctly adjusted system operating on 1380 kHz. To avoid the misleading effects of seasonal conductivity changes that might appear to distort measured directional antenna patterns in size or shape, the FCC requires that all the field strength measurements in a directional antenna proof of performance be made under "similar environmental conditions."

Standard Patterns

Theoretical (calculated) patterns can have nulls wherein the radiation at specific azimuths goes completely to zero. In practice, it is not possible to prove by field strength measurements that a null exists. Reradiation and scatter from objects external to the array limit the depth to which a pattern minimum can be proven. Additionally, operational variations in phase and ratio parameters will increase radiation in any direction where the deepest possible minimum has been previously established. To accommodate these limitations, the FCC authorizes a *standard pattern* for each directional antenna station. Standard patterns exceed the theoretical pattern at all azimuths by specified and easily calculated amounts. It is required that the radiation from a directional station not exceed its standard pattern. All U.S. stations employing directional antennas have FCC specified standard patterns. These supersede all earlier patterns based on theoretical calculations or on field strength measurements. The standard pattern radiation values are now used exclusively in all calculations of coverage and interference.

Augmented Patterns

Augmentation is applied to the standard pattern when the measured field strength is exceeded in discrete directions but does not cause interference to other stations. When augmentation is desired, it is achieved by applying Equation (C-9) in Appendix C.

REFERENCES

The first 12 references listed were published by Carl E. Smith Electronics, Inc.

1. *Directional Antenna Patterns.*
2. *Theory and Design of Directional Antennas.*
3. *Standard Broadcast Antenna Systems.*
4. *Design and Operation of Directional Antennas.*
5. *Directional Antenna Pattern Shapes.*
6. *Radiation Characteristics of Transmitting Antennae* (An Introduction to Directional Antenna Pattern Design.)
7. *Directional Antenna Design Example.*
8. *Parasitic Reradiation.*
9. *Introduction to Directional Antenna Systems.*
10. *Instructions for Installation of Radio Broadcast Stations Ground Systems.*
11. *Log Periodic Antenna Design Handbook.*
12. *Radio Broadcast Ground Systems.*
13. Ballantine, Stuart., "On the Optimum Transmitting Wavelength for a Vertical Antenna Over Perfect Earth," Proceedings of I.R.E. Vol. 12, pp. 833–839, December, 1924.
14. Brown, George H., "A Critical Study of the Characteristics of Broadcast Antennas as Affected by Current Distribution," Proceedings of I.R.E., Vol. 24, pp. 48–81, January, 1936.
15. Harmon, Ralph N., "Some Comments on Broadcast Antennas," Proceedings of I.R.E., Vol. 24, pp. 36–47, January, 1936.
16. Jeffers, C. L., "An Antenna for Controlling the Nonfading Range of Broadcasting Stations," Proceedings IRE, Vol. 36. pp. 1426–1431, November, 1948.
17. Morrison, J. F., and P. E. Smith, "The Shunt Excited Tower," Proceedings of IRE, vol. 25, pp. 673–696, June, 1937.
18. Smith, Carl E., and Johnson, Earl M., "Performance of Short Antennas," Proceedings of I.R.E., Vol. 35, pp. 1026–1038, October 1947.

19. Smith, Carl E., Hall, John R. and Weldon, James O., "Very High-Power Long-Wave Broadcasting Antennas," Proceedings of I.R.E., Vol. 42, No. 8, pp. 1222–1235, August, 1954.
20. Smith, Carl E., "A Critical Study of Several Antennas Designed to Increase the Primary Coverage of a Radio Broadcasting Transmitter," Professional Thesis at Ohio State University, Columbus, Ohio.
21. Smith, Carl E., "A Critical Study of Two Broadcast Antennas," Proceedings of I.R.E., Vol. 24, pp. 1329–1341, October, 1936.
22. Smith, Carl E., D. B. Hutton, and W. G. Hutton, "Performance of Sectionalized Broadcasting Towers," IRE Trans., pp. 22–34, December, 1955.
23. Smith, Carl E., "Short Low Loss Antennas," IEEE Transactions on Broadcasting, Vol. 35, No. 2, pp. 237–249, June, 1989.

APPENDIX A
DIRECTIONAL ANTENNAS FOR PATTERN SHAPE

Space Configuration

The Plan configuration of the k^{th} tower in an array is shown in Figure 4.8-A-1. A space view of the k^{th} tower and observation point P is shown in Figure 4.8-A-2.

Vector Diagram

The field strength at the point P in space for the k^{th} tower is shown in Figure 4.8-A-3. The space phasing in the horizontal plane is shown in Figure 4.8-A-4 and in the elevation plane the space phasing is reduced further as shown in Figure 4.8-A-5.

Generalized Equation

The vector equation to express the vectors in Figure 4.8-A-6 is the generalized equation that can be used to express the pattern shape for a directional antenna array of n towers. The equation in condensed form is,

$$E = \sum_{k=1}^{k=n} E_k f_k(\Theta)\beta_k \qquad \text{[A-1]}$$

where:

E = the total effective field strength vector at unit distance (P) for the antenna array with respect to the voltage vector reference axis. This vector makes the angle β with respect to this axis as shown in Figure 4.8-A-6.
k = the k^{th} tower in the directional antenna system
n = the total number of towers in the directional antenna array
E_k = the magnitude of the field strength at unit distance in the horizontal plane produced by the k^{th} tower acting alone
$f_k(\Theta)$ = vertical radiation characteristic of the k^{th} antenna as given in Equation A-3
Θ = elevation angle of the observation point P measured up from the horizon in degrees

$$\beta_k = S_k \cos\Theta \cos(\phi_k - \phi) + \psi_k \qquad \text{[A-2]}$$

is the phase relation of the field strength at the observation point P for the k^{th} tower taken with respect to the voltage vector reference axis

$$S_k \cos(\phi_k - \phi)\cos\Theta$$

is the space phasing portion of β_k due to the location of the k^{th} tower and ψ_k is the phasing portion of β_k

S_k = electrical length of spacing of the k^{th} tower in the horizontal plane from the space reference point
ϕ_k = true horizontal azimuth orientation of k^{th} tower with respect to the space reference axis
ϕ = true horizontal azimuth angle of the direction to the observation point P (measured clockwise from true north)
ψ_k = time phasing portion of β_k due to the electrical phase angle of the voltage (or current) in the k^{th} tower taken with respect to the voltage vector reference axis

The shape of any directional antenna pattern can be computed by applying the above equations, however, many directional antenna arrays can be designed by simplified versions of this equation.

For a vertical antenna having a sinusoidal current distribution with a current node at the top, the vertical radiation characteristic takes on the form

$$f(\Theta) = \frac{\cos(G\sin\Theta) - \cos G}{(1 - \cos G)\cos\Theta} \qquad \text{[A-3]}$$

where

$f(\Theta)$ = vertical radiation characteristic
G = electrical height of the antenna in electrical degrees
Θ = elevation angle of the observation point measured up from the horizon in degrees

The vertical radiation characteristics in Equation A-3 are graphed in Figure 4.8-A-7.

For a top-loaded tower the formula is:

$$f(\Theta) = \frac{\cos B\cos(A\sin\Theta) - \cos G - \sin B\sin\Theta\sin(A\sin\Theta)}{\cos\Theta(\cos B - \cos G)}$$

This is the vertical radiation characteristic for a top-loaded tower of height A and top-loaded to a height of $G = A + B$.

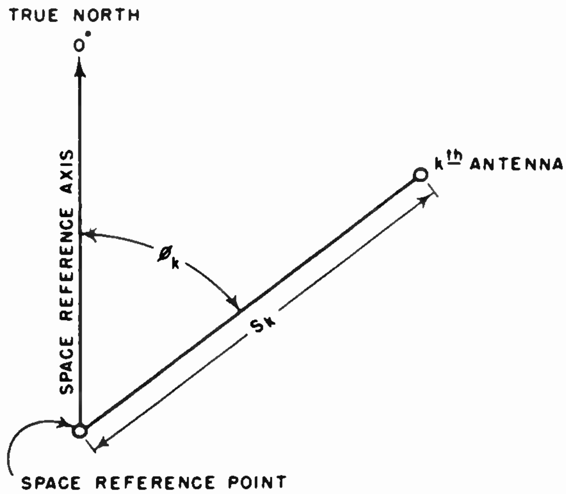


Figure 4.8-A-1. Plan view of space configuration of k^{th} antenna.

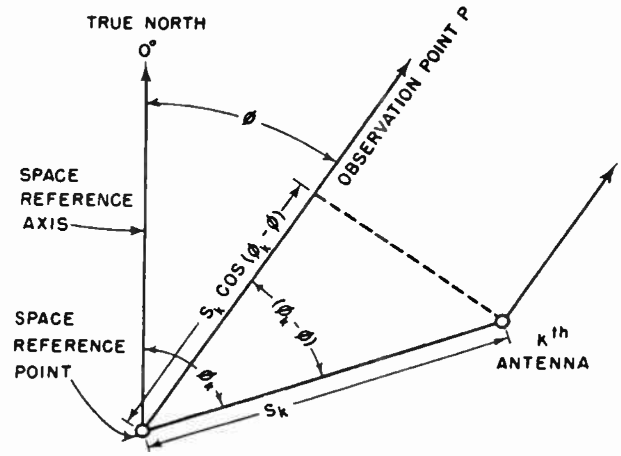


Figure 4.8-A-4. Plan view of k^{th} antenna showing space phasing in the horizontal plane.

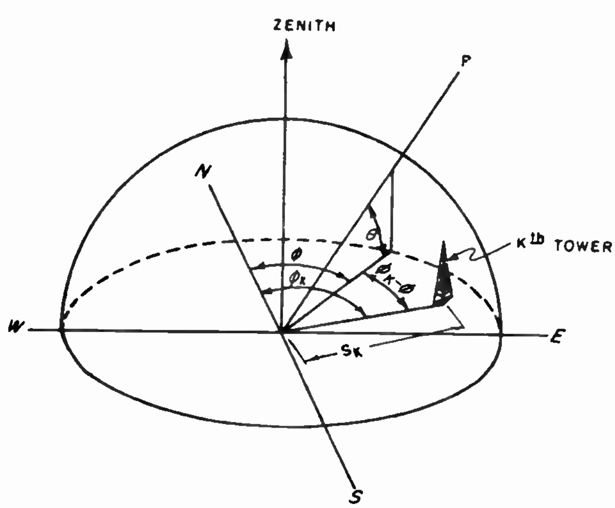


Figure 4.8-A-2. Space view of observation point P and the k^{th}.

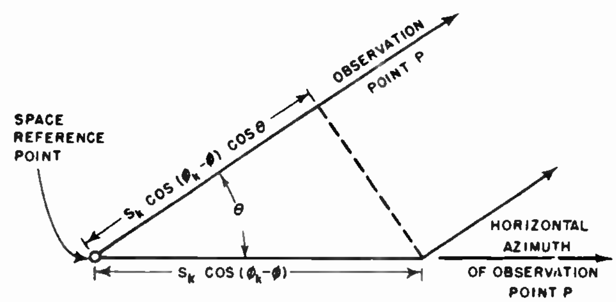


Figure 4.8-A-5. Elevation angle θ shortens the spacing S^k to the value of $S^k \cos\theta$.

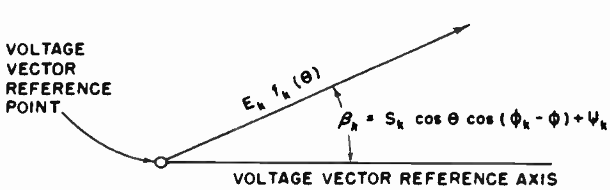


Figure 4.8-A-3. Voltage vector diagram for the k^{th} antenna.

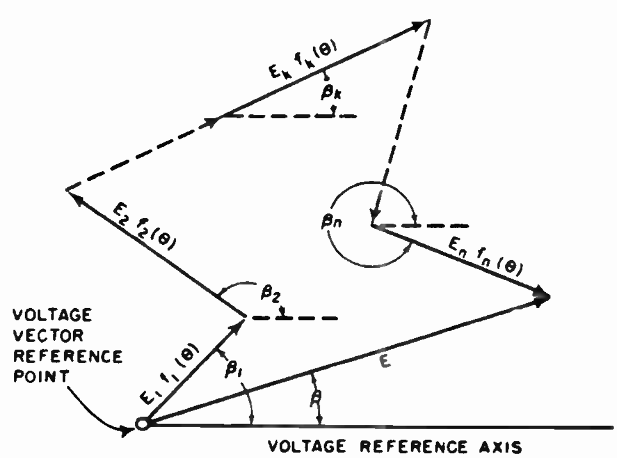


Figure 4.8-A-6. Summation of field strength vectors for n antennas in the directional antenna array.

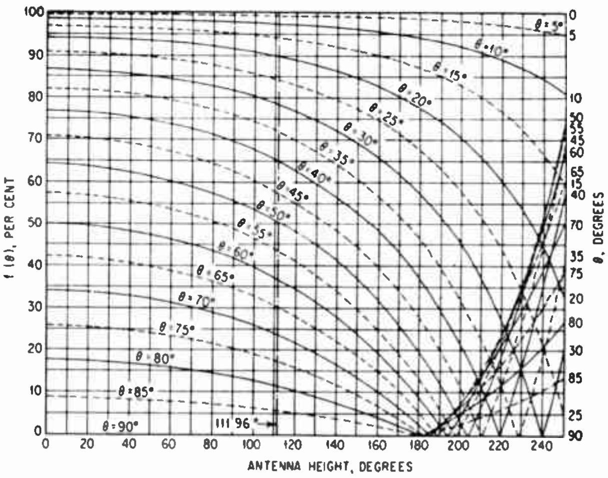


Figure 4.8-A-7. Vertical-radiation characteristics as a function of electrical tower height for various values of elevation angle.

For a two section top-loaded tower as shown in Figure 4.8-5, the formula is:

$$\cos B \cos(A \sin \Theta) - \cos G + \frac{\sin B \cos(H - C)\,\cos(C \sin \Theta)}{\sin(H - A)}$$

$$f(\Theta) = \frac{-\dfrac{\sin B \sin \Theta \sin (H - C) \sin(C \sin \Theta)}{\sin(H - A)} - \dfrac{\sin B \cos (H - A) \cos(A - \Theta)}{\sin(H - A)}}{\cos\{\cos B - \cos G + [\sin B/\sin(H - A)]\,(\cos \overline{H - C} \cos \overline{H - A})\}} \qquad \text{[A-4]}$$

This is the vertical radiation characteristic equation for a two-section sectionalized tower. The same procedure can be applied if more than two sections are involved, such as shown in Figure 4.8-6.

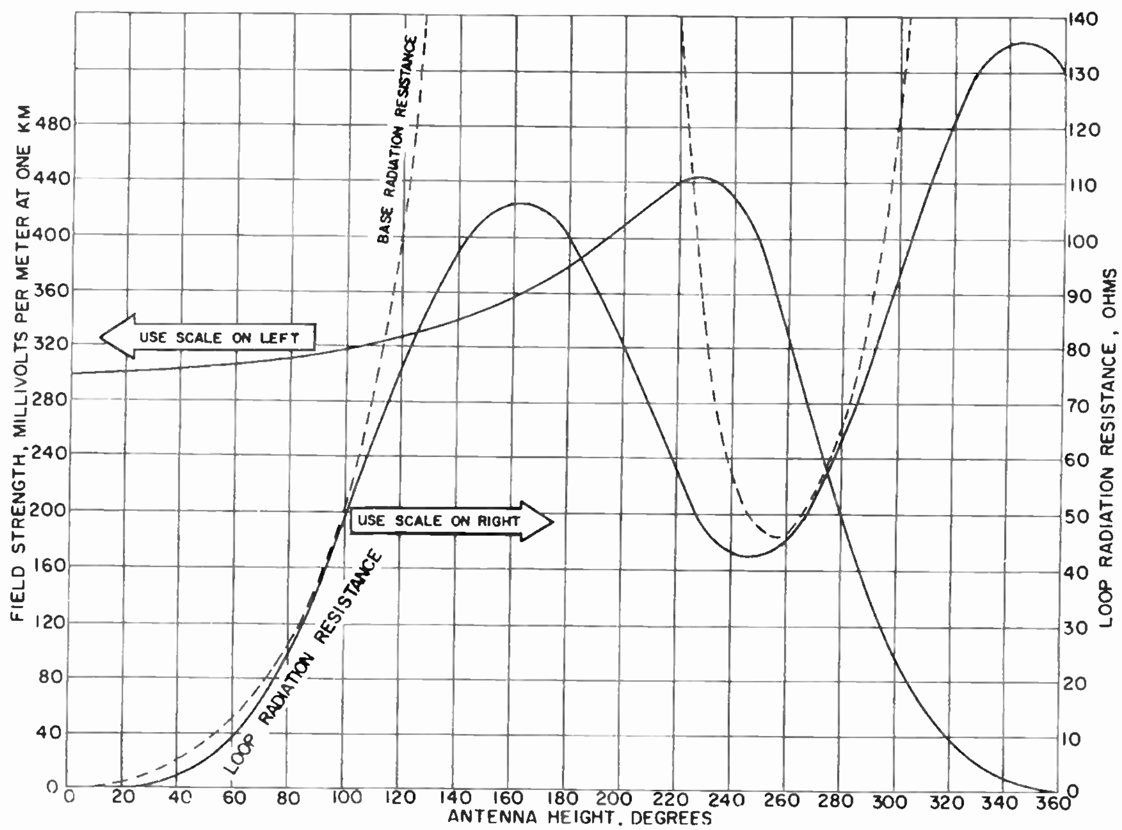


Figure 4.8-A-8. Inverse field strength at 1 km for 1 kW, loop and base radiation resistance as a function of tower height over a perfectly conducting earth.

Theoretical Self-Loop and Base-Radiation Resistance

It is useful to know the theoretical loop and base resistance of a vertical radiator. This information is presented graphically in Figure 4.8-A-8 along with the theoretical inverse field strength at 1 km.

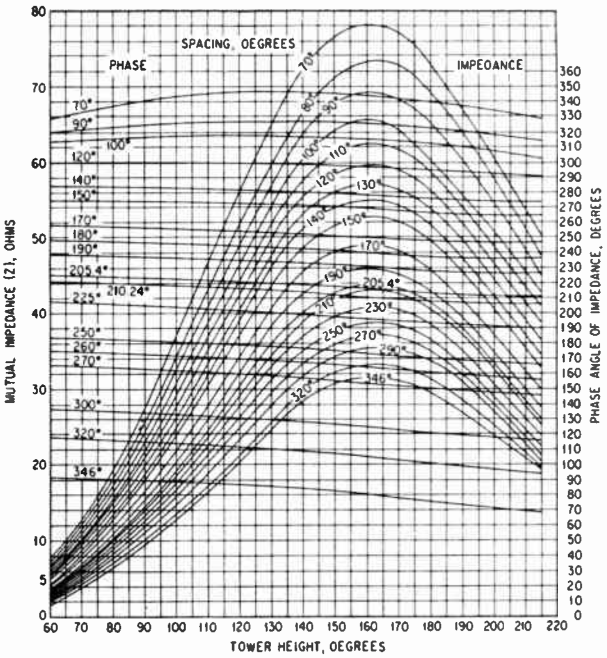


Figure 4.8-A-9. Loop mutual impedance and phase angle between two towers of equal height.

Mutual Impedance Curves

The value of mutual impedance for most tower heights and spacing is given in Figure 4.8-A-9. The loop mutual impedance between quarter-wave towers is shown in Figure 4.8-A-10.

Horizontal RMS Field Strength

The field strength gain or loss of a two-tower array for various values of phasing and spacing is shown in Figure 4.8-A-11.

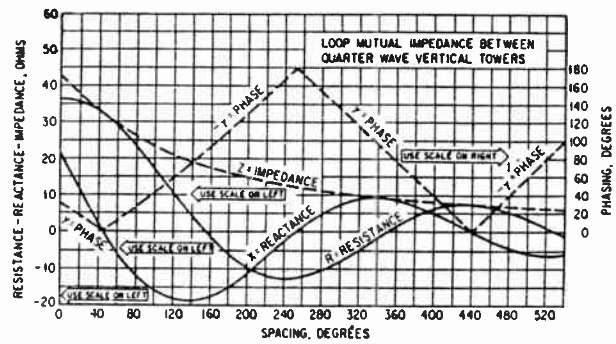


Figure 4.8-A-10. Loop mutual impedance between quarter-wave vertical towers.

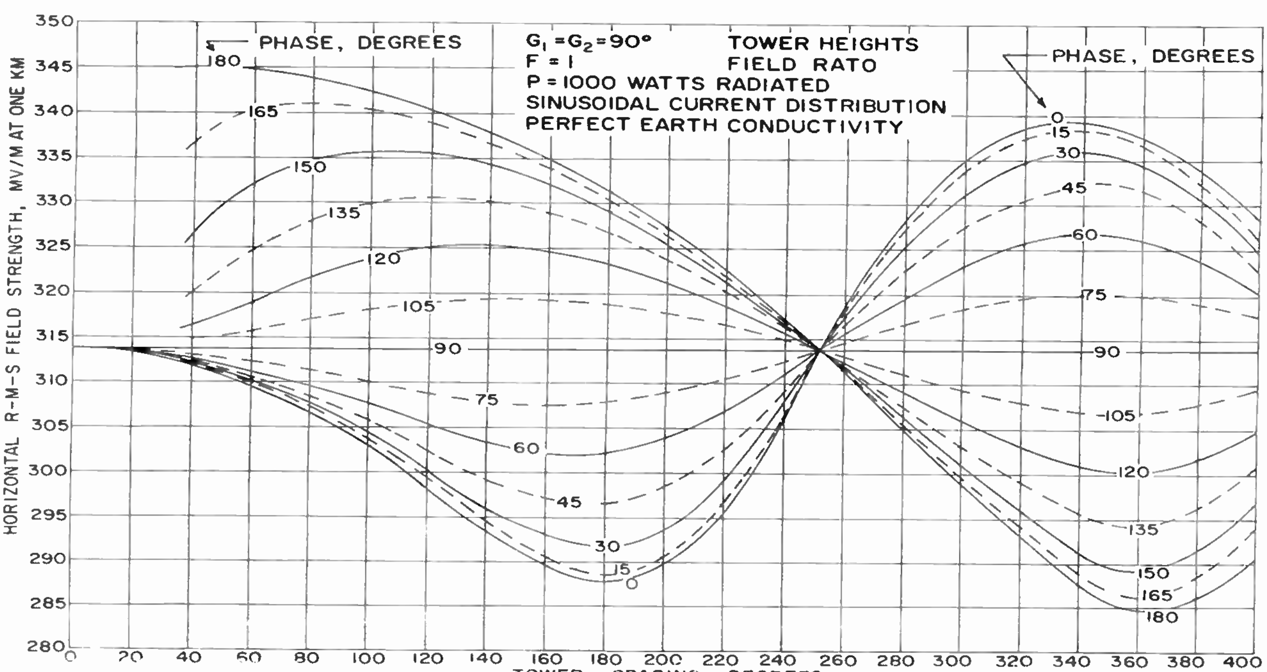


Figure 4.8-A-11. Horizontal RMS field strength of two-tower directional antenna.

APPENDIX B
SYSTEMATIZATION OF TWO TOWER PATTERNS

The pattern numbering system has been devised to furnish the antenna parameters in an orderly fashion as described at the bottom of each Figure. To simplify the numbering system, the spacing and phasing of each antenna is shifted in steps of 45°. The spacing varies from 45° in Figures 4.8-B-1 to 1,440 in Figure 4.8-B-4 while the phasing varies from 0° to 315° in each figure. The field strength of each antenna, E1 and E2, is expressed in percent of the maximum lobe field strength (50%) of the horizontal pattern. Finally, the last two digits specify, in percent, the RMS field strength in the horizontal plane of the two tower directional antenna array.

The detailed patterns which follow are for spacings in steps of 15° up to 360° degrees. The phasings are only presented from 0° to 180° since the same patterns, turned 180°, give the results for phasings from 180° to 360°.

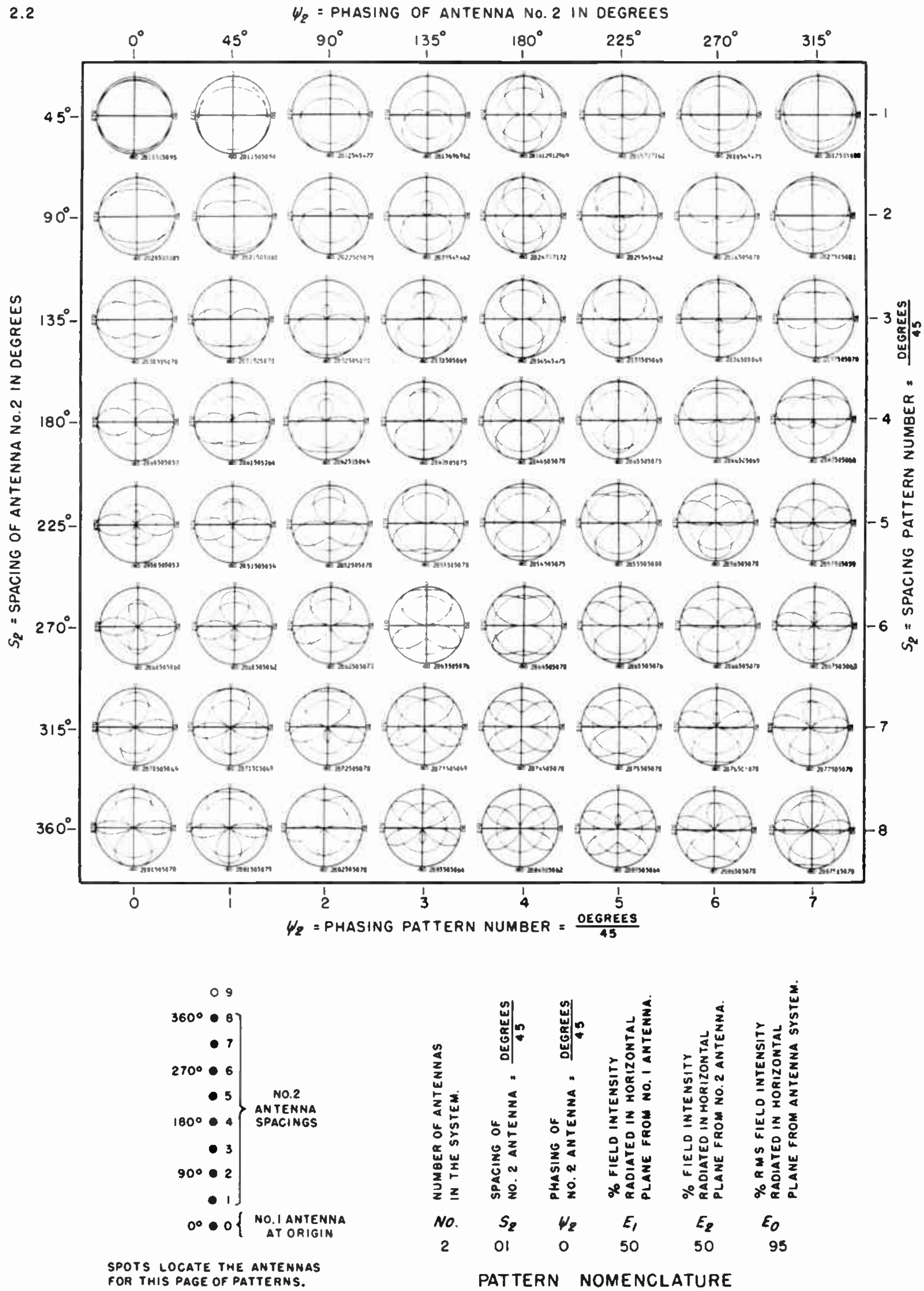
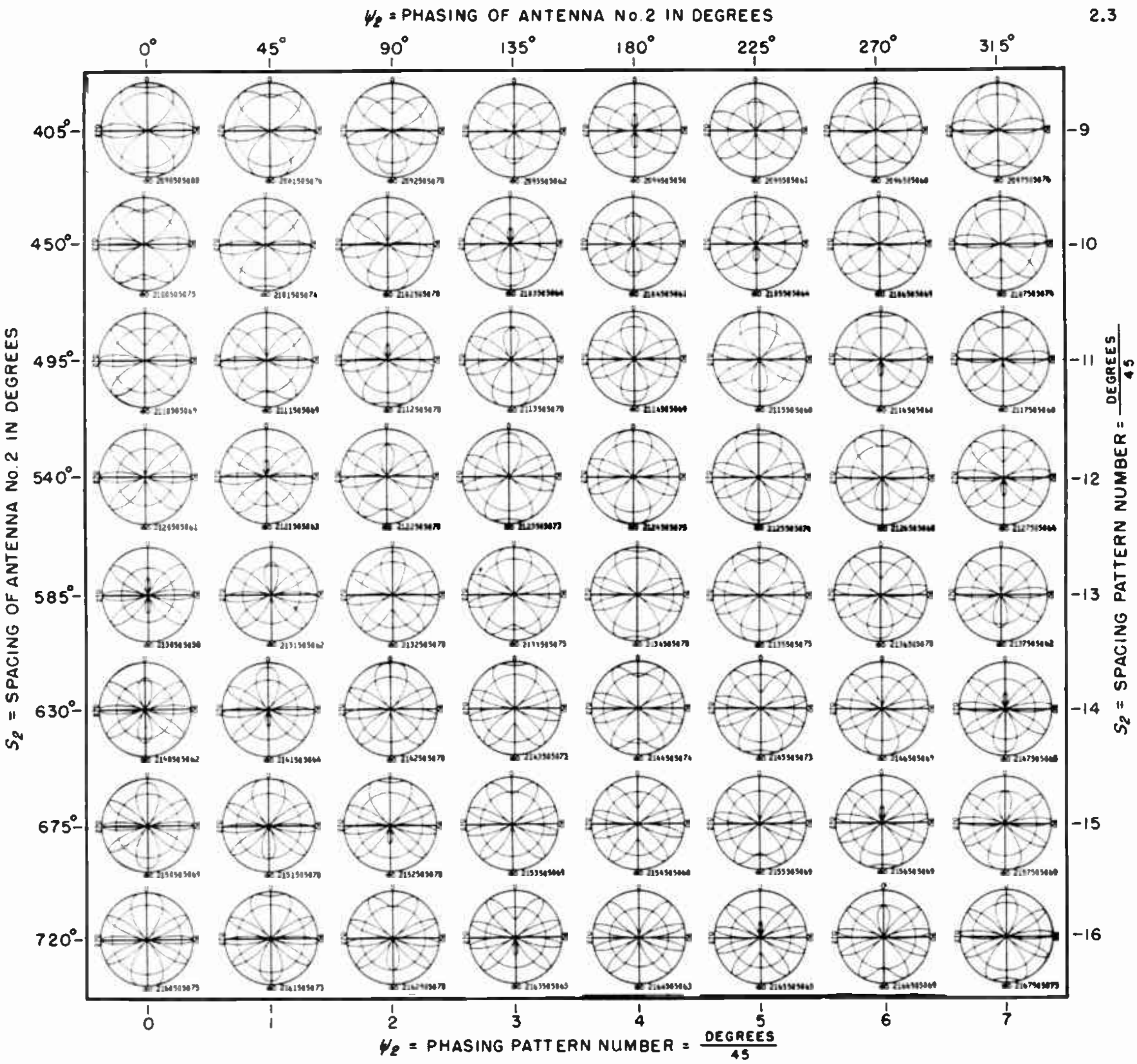
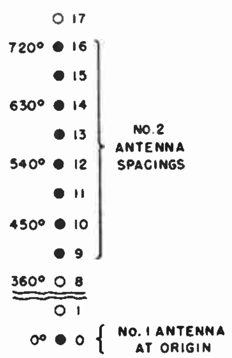


Figure 4.8-B-1. Effect of changing Phase and Spacing on Directional Antenna Patterns.





SPOTS LOCATE THE ANTENNAS FOR THIS PAGE OF PATTERNS.

NUMBER OF ANTENNAS IN THE SYSTEM.	SPACING OF NO. 2 ANTENNA = $\frac{\text{DEGREES}}{45}$	PHASING OF NO. 2 ANTENNA = $\frac{\text{DEGREES}}{45}$	% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 1 ANTENNA.	% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 2 ANTENNA.	% RMS FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM ANTENNA SYSTEM.
NO.	S_2	ψ_2	E_1	E_2	E_0
2	09	0	50	50	80

PATTERN NOMENCLATURE

Figure 4.8-B-2.

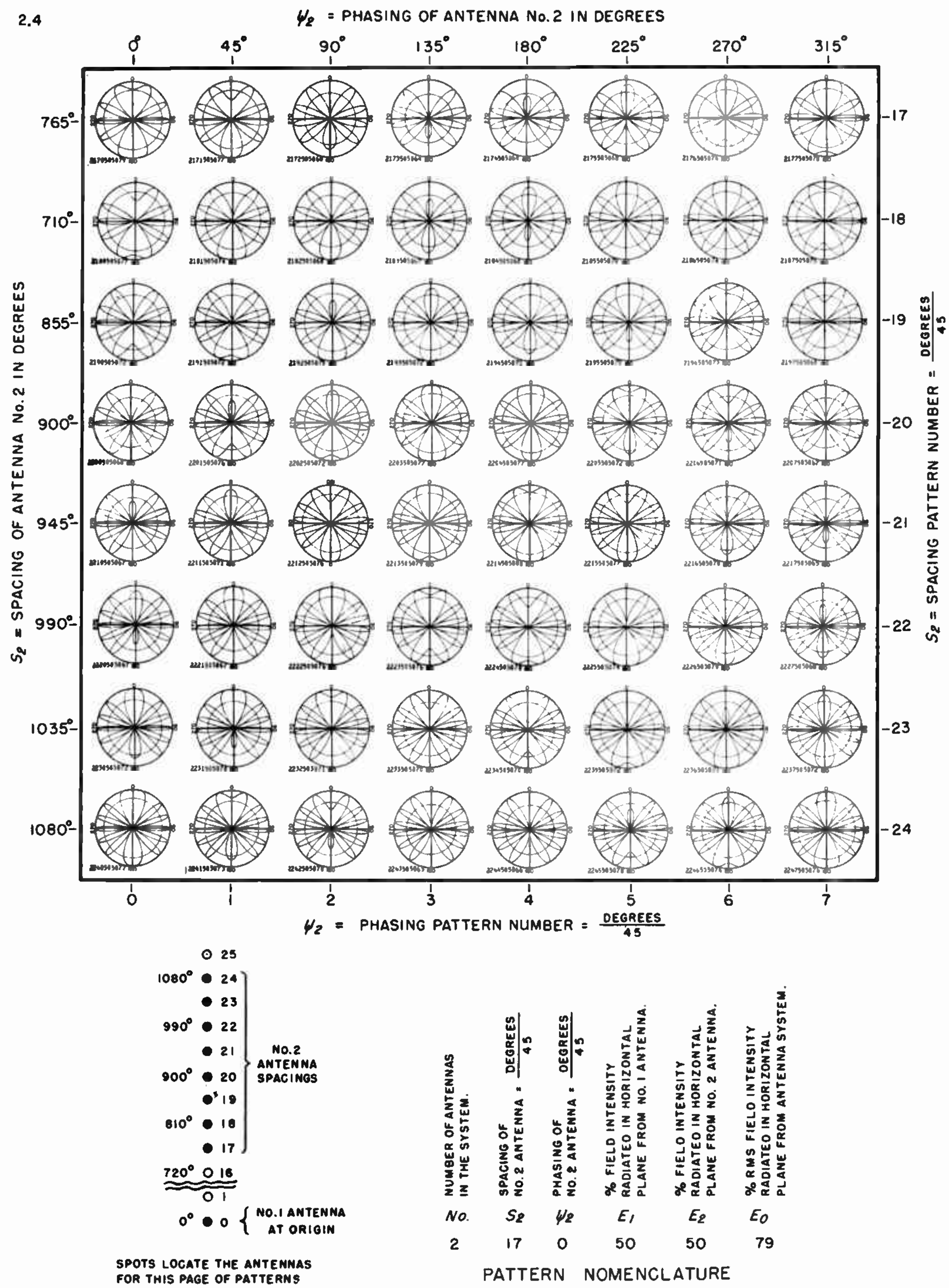

2.4
ψ_2 = PHASING OF ANTENNA No. 2 IN DEGREES
0° 45° 90° 135° 180° 225° 270° 315°
S_2 = SPACING OF ANTENNA No. 2 IN DEGREES
765° 710° 855° 900° 945° 990° 1035° 1080°
S_2 = SPACING PATTERN NUMBER = $\frac{\text{DEGREES}}{45}$
17 18 19 20 21 22 23 24
0 1 2 3 4 5 6 7
ψ_2 = PHASING PATTERN NUMBER = $\frac{\text{DEGREES}}{45}$
25
1080° 24
23
990° 22
21
900° 20
19
610° 18
17
NO. 2 ANTENNA SPACINGS
720° 16
1
0° 0
NO. 1 ANTENNA AT ORIGIN
SPOTS LOCATE THE ANTENNAS FOR THIS PAGE OF PATTERNS
NUMBER OF ANTENNAS IN THE SYSTEM.
SPACING OF NO. 2 ANTENNA = $\frac{\text{DEGREES}}{45}$
PHASING OF NO. 2 ANTENNA = $\frac{\text{DEGREES}}{45}$
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 1 ANTENNA.
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 2 ANTENNA.
% RMS FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM ANTENNA SYSTEM.
No. S_2 ψ_2 E_1 E_2 E_0
2 17 0 50 50 79
PATTERN NOMENCLATURE

Figure 4.8-B-3.

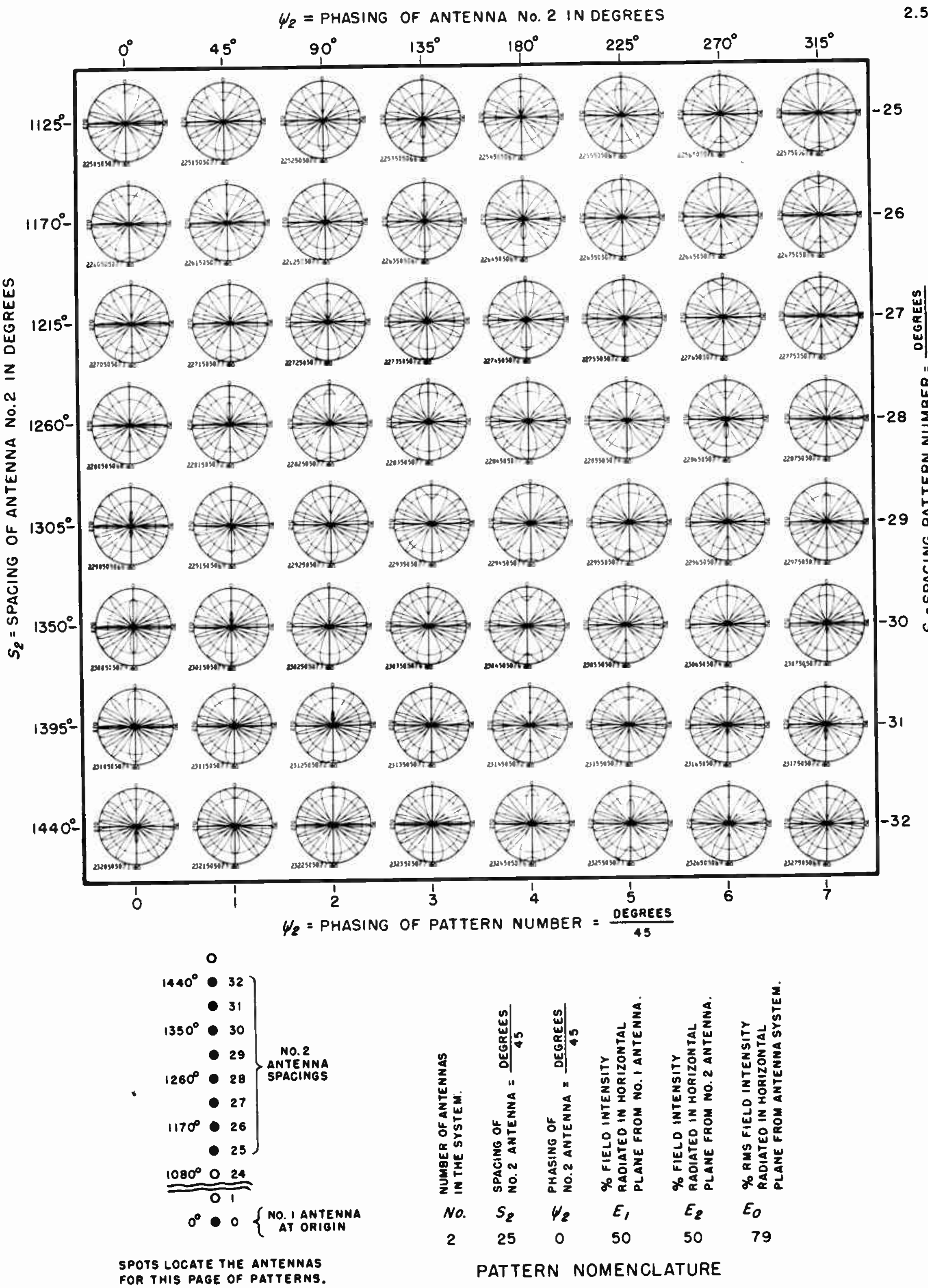

ψ_2 = PHASING OF ANTENNA No. 2 IN DEGREES
2.5
0° 45° 90° 135° 180° 225° 270° 315°
1125° 1170° 1215° 1260° 1305° 1350° 1395° 1440°
S_2 = SPACING OF ANTENNA No.2 IN DEGREES
-25 -26 -27 -28 -29 -30 -31 -32
S_2 = SPACING PATTERN NUMBER = DEGREES / 45
0 1 2 3 4 5 6 7
ψ_2 = PHASING OF PATTERN NUMBER = DEGREES / 45
0
1440° 32
31
1350° 30
29
1260° 28
27
1170° 26
25
NO. 2 ANTENNA SPACINGS
1080° 24
1
0° 0
NO. 1 ANTENNA AT ORIGIN
SPOTS LOCATE THE ANTENNAS FOR THIS PAGE OF PATTERNS.
NUMBER OF ANTENNAS IN THE SYSTEM.
SPACING OF NO. 2 ANTENNA = DEGREES / 45
PHASING OF NO. 2 ANTENNA = DEGREES / 45
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 1 ANTENNA.
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 2 ANTENNA.
% RMS FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM ANTENNA SYSTEM.
No. S_2 ψ_2 E_1 E_2 E_0
2 25 0 50 50 79
PATTERN NOMENCLATURE

Figure 4.8-B-4.

2.6

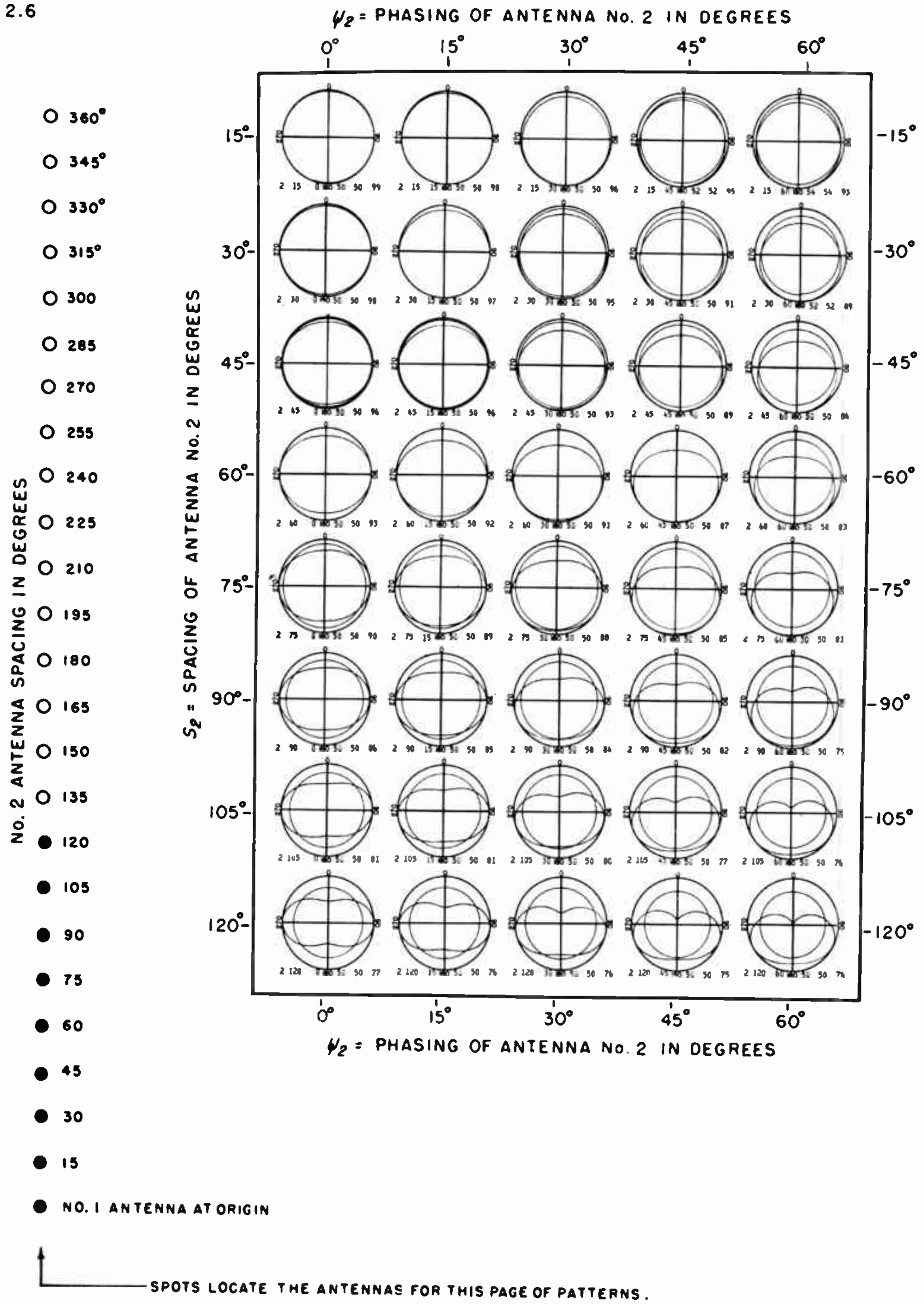
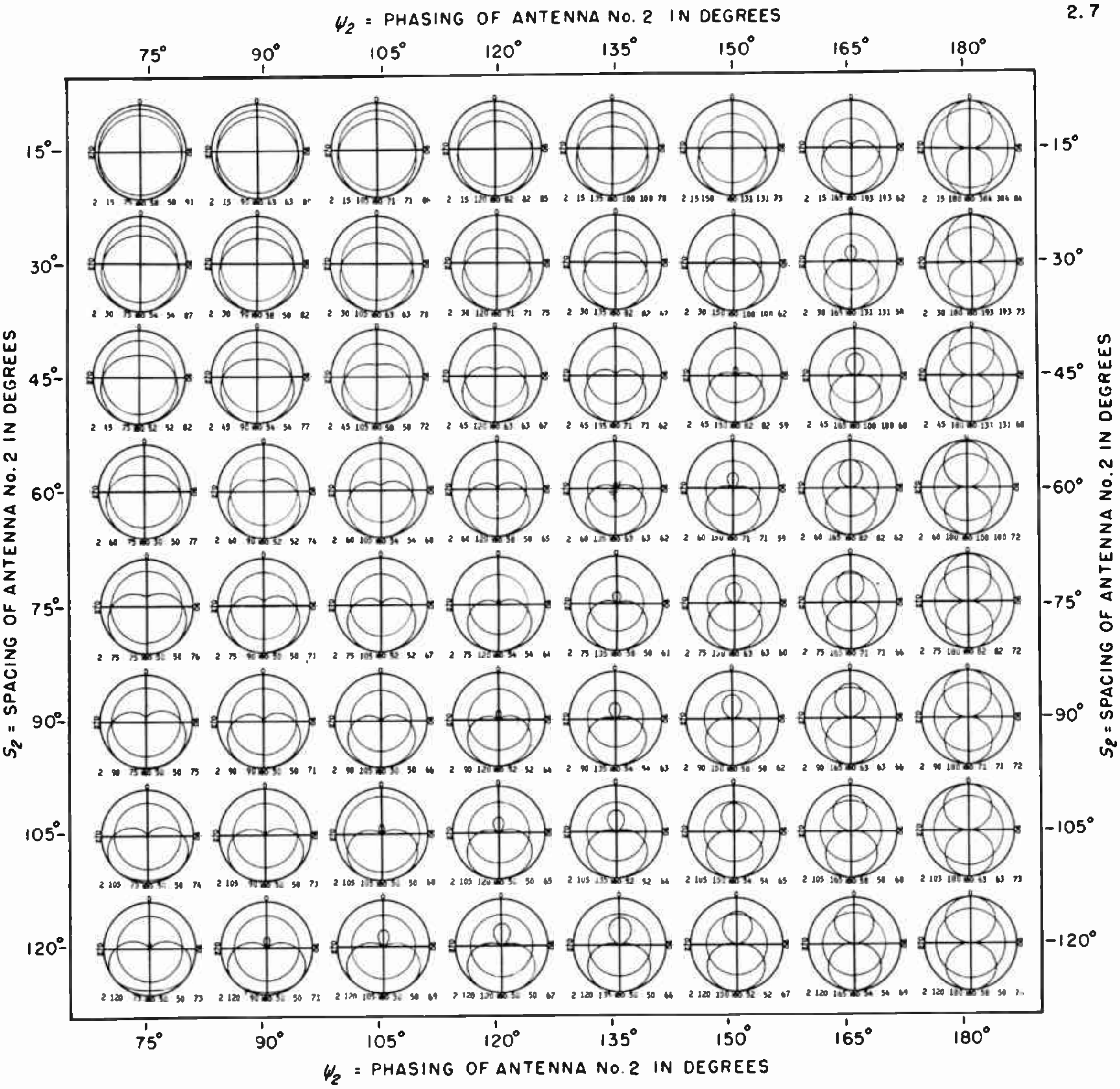


Figure 4.8-B-5.

2.7



NUMBER OF ANTENNAS IN THE SYSTEM	SPACING OF No. 2 ANTENNA IN DEGREES	PHASING OF No. 2 ANTENNA IN DEGREES	% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM No. 1 ANTENNA	% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM No. 2 ANTENNA	% RMS FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM ANTENNA SYSTEM
No.	S_2	ψ_2	E_1	E_2	E_0
2	15	75	58	58	91

PATTERN NOMENCLATURE

Figure 4.8-B-6.

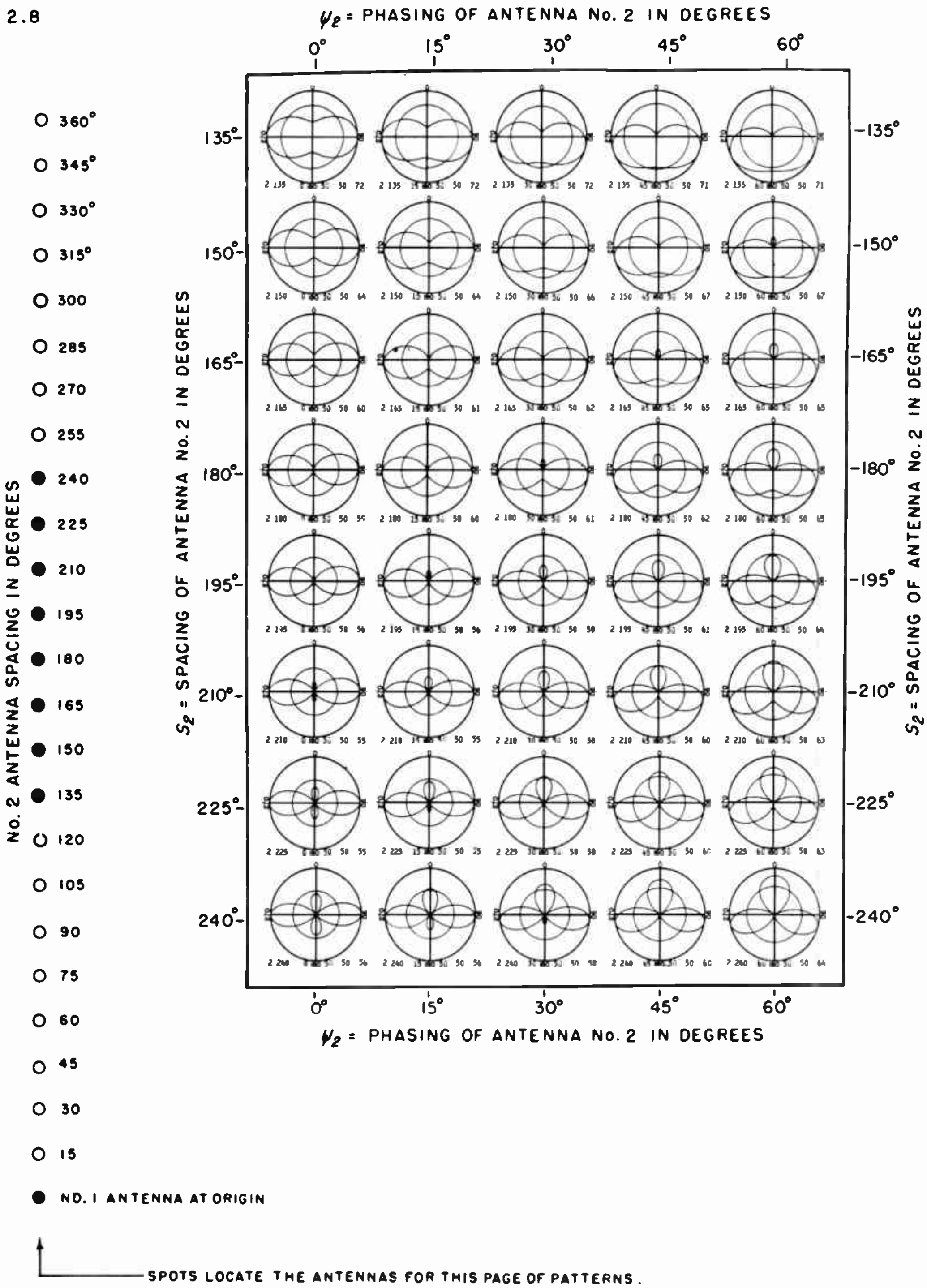

2.8
ψ2 = PHASING OF ANTENNA No. 2 IN DEGREES
0°
15°
30°
45°
60°
135°
150°
165°
180°
195°
210°
225°
240°
-135°
-150°
-165°
-180°
-195°
-210°
-225°
-240°
S2 = SPACING OF ANTENNA No. 2 IN DEGREES
NO. 2 ANTENNA SPACING IN DEGREES
360°
345°
330°
315°
300
285
270
255
240
225
210
195
180
165
150
135
120
105
90
75
60
45
30
15
NO. 1 ANTENNA AT ORIGIN
SPOTS LOCATE THE ANTENNAS FOR THIS PAGE OF PATTERNS.

Figure 4.8-B-7.

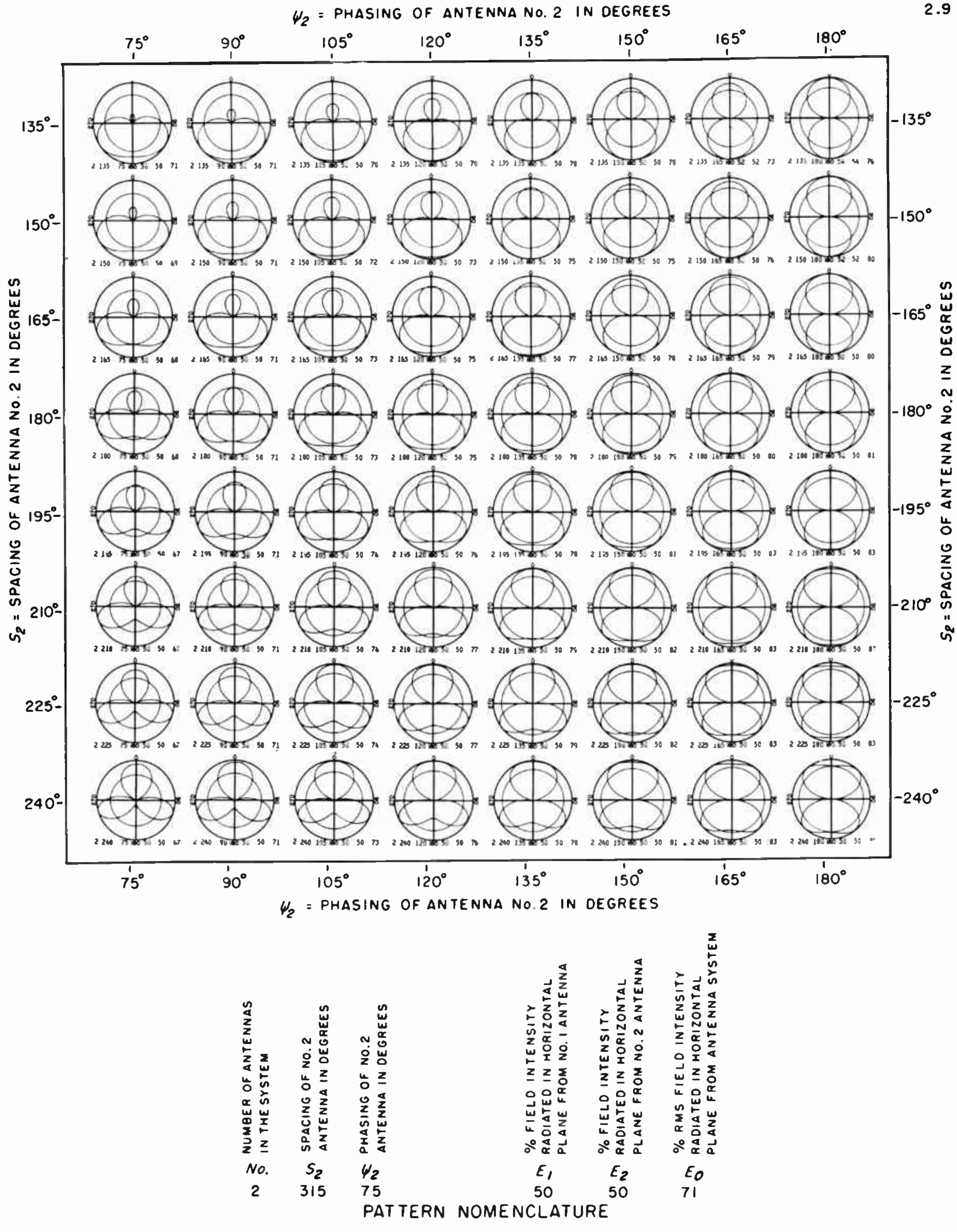

2.9
ψ_2 = PHASING OF ANTENNA No. 2 IN DEGREES
75° 90° 105° 120° 135° 150° 165° 180°
S_2 = SPACING OF ANTENNA No. 2 IN DEGREES
135° 150° 165° 180° 195° 210° 225° 240°
NUMBER OF ANTENNAS IN THE SYSTEM
SPACING OF NO. 2 ANTENNA IN DEGREES
PHASING OF NO. 2 ANTENNA IN DEGREES
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 1 ANTENNA
% FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM NO. 2 ANTENNA
% RMS FIELD INTENSITY RADIATED IN HORIZONTAL PLANE FROM ANTENNA SYSTEM
No. S_2 ψ_2 E_1 E_2 E_0
2 315 75 50 50 71
PATTERN NOMENCLATURE

Figure 4.8-B-8.

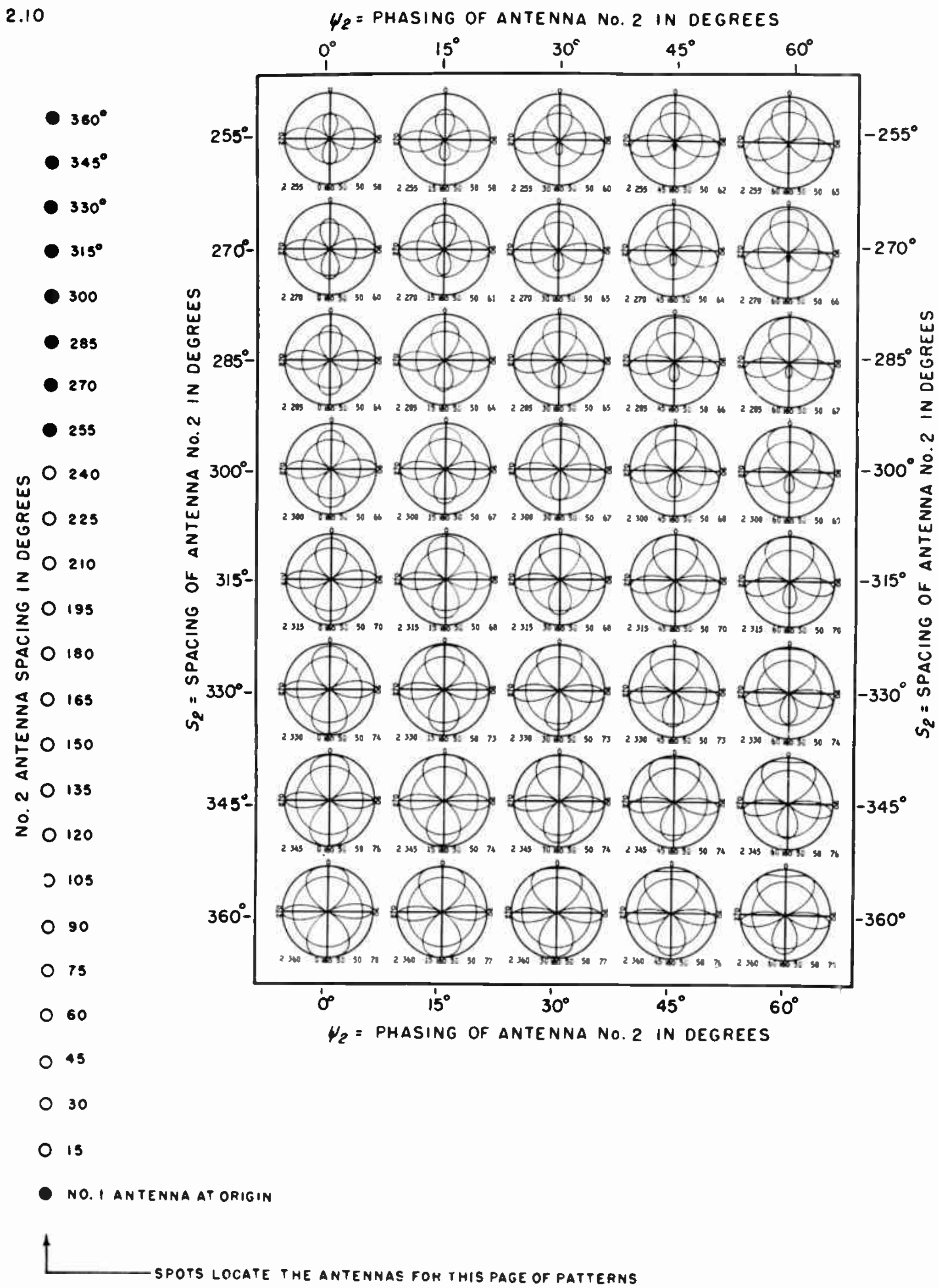

2.10
ψ_2 = PHASING OF ANTENNA No. 2 IN DEGREES
0° 15° 30° 45° 60°
S_2 = SPACING OF ANTENNA No. 2 IN DEGREES
255° 270° 285° 300° 315° 330° 345° 360°
-255° -270° -285° -300° -315° -330° -345° -360°
No. 2 ANTENNA SPACING IN DEGREES
360° 345° 330° 315° 300 285 270 255 240 225 210 195 180 165 150 135 120 105 90 75 60 45 30 15
NO. 1 ANTENNA AT ORIGIN
SPOTS LOCATE THE ANTENNAS FOR THIS PAGE OF PATTERNS

Figure 4.8-B-9.

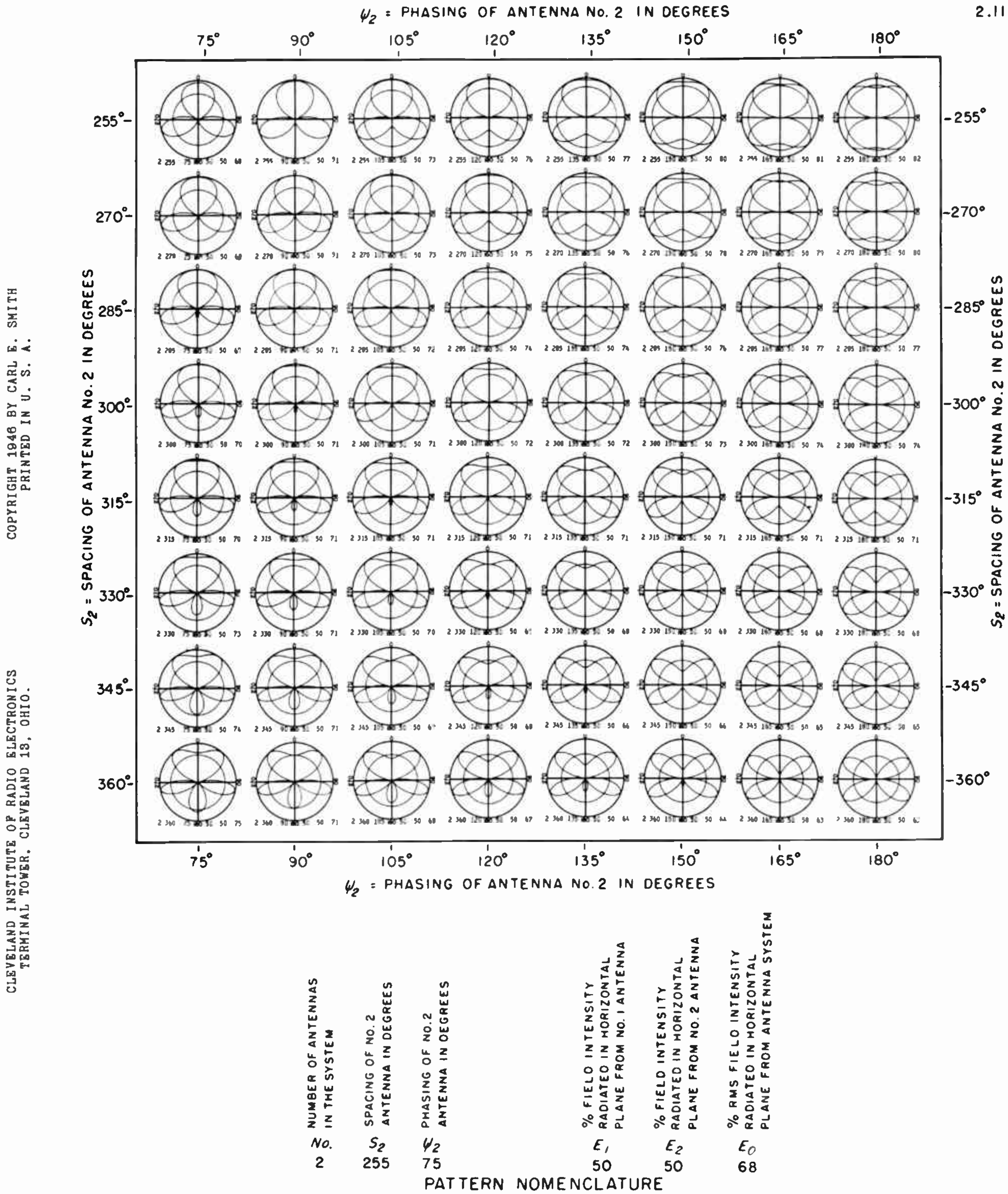


Figure 4.8-B-10.

APPENDIX C
PATTERN DEVELOPMENT OF DIRECTIONAL ANTENNAS

Theoretical Pattern Equation

The theoretical pattern equation of Appendix A can be written as follows by changing the k^{th} tower to the i^{th} tower to conform with the FCC practice, thus:

$$E(\phi, \Theta)_{th} = \left| k \sum_{i=1}^{n} F_i(\Theta) \underline{/ S_i \cos\Theta \cos(\Phi_i - \Phi) + \Psi_i} \right| \quad [C-1]$$

where:

k = multiplying constant which determines pattern size

Standard Pattern Equation

The standard pattern equation is obtained from Equation C-1 by adding the quadrature term Q to fill minimums and increase the size by 5%, thus:

$$E(\Phi, \Theta)_{std} = 1.05\sqrt{[E(\Phi, \Theta)_{th}]^2 + Q^2} \quad [C-2]$$

where Q is the greater of the quantities:

$$0.025g(\Theta)E_{rss} \quad [C-3]$$

or

$$10.0g(\Theta)\sqrt{P_{kw}} \quad [C-4]$$

where $g(\Theta)$ is the vertical plane distribution factor, $f(\Theta)$, for the shortest element in the array (see Equation C-2 above; also see FCC Rules Section 73.190, Figure 5).

If the shortest element has an electrical height in excess of 0.5 wavelength, $g(\Theta)$ shall be computed as follows:

$$g(\Theta) = \frac{\sqrt{\{f(\Theta)\} + 0.0625}}{1.030776} \quad [C-5]$$

$$E_{rss} = \sqrt{\sum_{i=1}^{n} E_i^2} \quad [C-6]$$

As an example, consider a two-tower array. The theoretical pattern equation (Equation C-1) becomes:

$$E = E_1 f_1(\Theta) \ \underline{/0^\circ} + E^2 f_2(\Theta) \ \underline{/ S_2 \cos\Theta \cos(\phi_2 - \phi) + \psi_2} \quad [C-7]$$

Now, for 5 kW, 90° towers, and the following parameters:

Tower number	Height G^0	Field Ratio	Spacing S^0	True Bearing ϕ^0	Phase ψ
1	90	1.0	0	0	0
2	90	1.0	90	0	−90

E_{rss} (theoretical pattern) = 691.92 mV/m
Q (quadrature term) = 21.60 mV/m
E_{rms} (standard pattern) = 726.87 mV/m

A plot of the theoretical and standard patterns is shown in Figure 4.8-C-1.

The minimum horizontal field strength (at one km) when the theoretical field strength goes to zero is given by Equation C-2 for a standard pattern along the ground using Equation C-4 with g(Θ) = 1.0. For 1 kW and under, Q is 6 according to FCC Rules. For various FCC licensed values of power, the minimum field strength values are as follows:

P_{kw}	Q	E_{mvm}
0.25	6.0	10.14
0.50	6.0	10.14
1.00	6.0	10.14
2.50	9.49	16.03
5.00	13.42	22.67
10.00	18.97	32.06
25.00	30.00	50.69
50.00	42.43	71.69

The minimum field strength (at 1 kW) for any elevation of a standard pattern, by Equation C-5, is:

$$g(\Theta) = \frac{\sqrt{0 + 0.0625}}{1.030776} = 0.2425$$

Augmented Pattern Equation

The augmented pattern equation is obtained by adding an augmentation quadrature term to the standard pattern as given here:

$$E_{(\phi,\Theta)aug} = \sqrt{\{E_{(\phi,\Theta)std}\}^2 + A\left\{g(\Theta)\cos\left(180\frac{D_A}{S}\right)\right\}} \quad [C-8]$$

where:

$E_{(\phi,\Theta)aug}$ = augmented pattern radiation value at azimuth, elevation
$E_{(\phi,\Theta)std}$ = standard pattern radiation value at azimuth, elevation
A = augmentation constant (insert equation C-10)

$$g(\Theta) = \frac{\sqrt{\{f(\Theta)\}^2 + 0.0625}}{1.030776} \quad [C-9]$$

S = span of augmentation in degrees
D_A = angular distance from center of span

The principle of augmentation is illustrated in the cardioid pattern of Figure 4.8-C-2.

The FCC has converted all augmented directional patterns to a table for each station as shown in the example of Figure 4.8-C-3. In this case there were six augmentations as tabulated in Figure 4.8-C-4 and shown on the polar chart of Figure 4.8-C-5.

It should be noted that where the spans overlap, Equation C-9 is applied repeatedly, once for each augmentation, proceeding clockwise from true north.

THEORETICAL vs STANDARD PATTERN

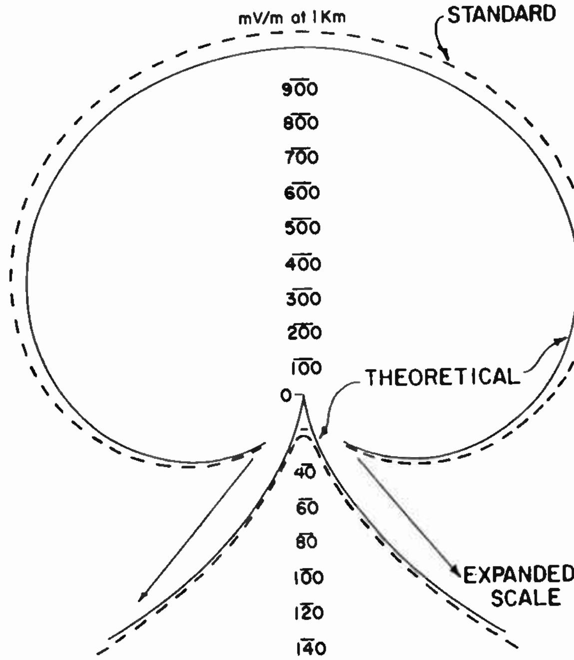


Figure 4.8-C-1. Theoretical and standard pattern.

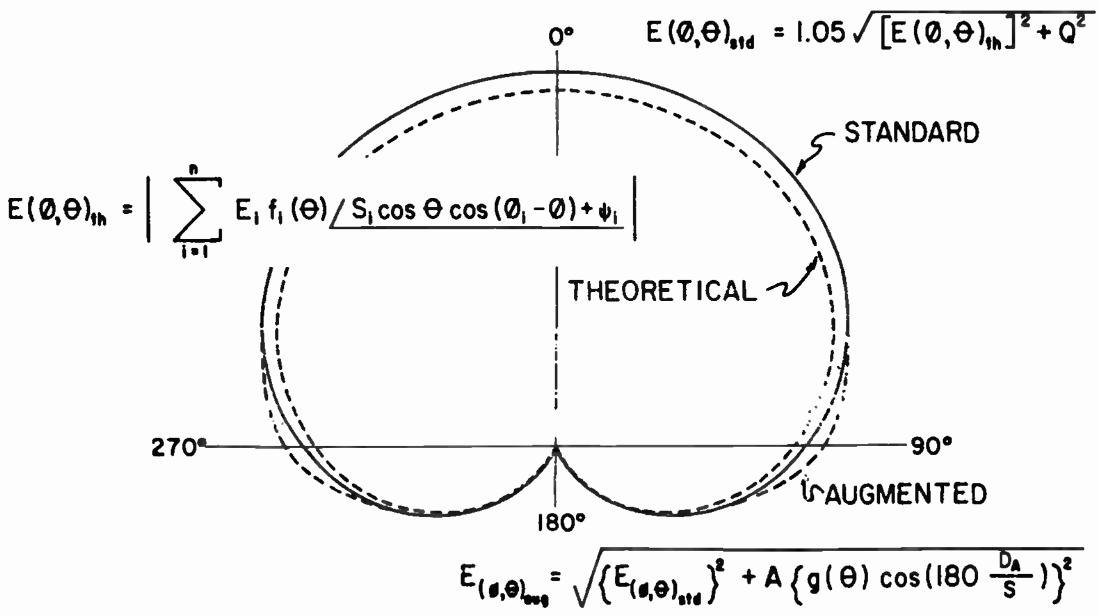


Figure 4.8-C-2. Theoretical, standard and augmented pattern.

Technical Parameters Resulting From Conversion of AM Broadcast Stations To Standard Patterns

STANDARD PATTERN CONVERSION NO.: 1280–22

FREQ. KHZ	CALL LETTER	CITY	STATE	PATTERN HRS.	PATTERN STATUS	CLASS
1280	WHVR	HANOVER	PA	N	LIC.	B

POWER KW	LATITUDE	LONGITUDE	PAT-MULT MV/M	TH-RMS MV/M	STD/AUG RMS-MV/M	PAT-RSS MV/M	Q-FACTOR
.500	39-49-11	77-00-25	131.27	143.00	150.54	185.65	6.0000

TOWER NO.	PHYS-HT (A)-DEG	TL-HT (B)-DEG	TOT-HT (C)-DEG	TL-HT (D)-DEG	FIELD RATIO	PHASE DEG.	SPACING DEG.	ORIENT DEG-TR	REF FLG
1	91.0	.0	.0	.0	1.000	149.5	.0	.0	
2	91.0	.0	.0	.0	1.000	.0	90.0	178.0	

AUGMENTATION DATA

CENTRAL AZIM. DEGREES TRUE	SPAN DEGREES	FIELD AT AZIM. MV/M
64.0	12.0	17.0
260.5	55.0	103.0
288.0	14.0	7.5
288.0	10.0	21.2
295.0	14.0	30.0
295.0	10.0	43.3

HORIZONTAL PLANE STANDARD/AUGMENTED RADIATION VALUES

AZ. DEG	FIELD MV/M	AZ. DEG	FIELD MV/M	AZ. DEG	FIELD MV/M	AZ. DEG	FIELD MV/M	AZ. DEG.	FIELD MV/M	AZ. DEG	FIELD MV/M
0	136.8	60	29.1	120	174.0	180	239.4	240	163.9	300	42.2
10	132.8	70	9.0	130	196.3	190	237.0	250	136.2	310	71.0
20	123.1	80	43.5	140	213.5	200	231.5	260	104.7	320	95.4
30	107.5	90	80.0	150	225.9	210	221.4	270	69.3	330	114.4
40	86.3	100	115.0	160	234.0	220	207.2	280	31.3	340	127.7
50	59.8	110	146.7	170	238.4	230	188.0	290	20.4	350	135.1

CONSTRUCTION PERMIT LIMITS

AZIMUTH DEG. TRUE	PRESENT MV/M	NEW MV/M
64.0	17.0	17.0
231.0	179.0	185.8
288.0	33.0	21.2
352.0	131.0	135.9

PATTERN MINIMA

AZIMUTH DEG. TRUE	FIELD MV/M
68.5	6.8
284.5	17.3
290.0	20.4
299.0	40.9

PATTERN MAXIMA

AZIMUTH DEG. TRUE	FIELD MV/M
178.0	239.4
288.0	21.2
296.0	44.2
358.0	136.9

Figure 4.8-C-3. FCC method of specifying augmentation.

CENTER AZIMUTH OF AUGMENTATION	SPAN DEGREES	EXTENT OF SPAN	FIELD AT CENTER SPAN AT 1 KM
64	12 ('6)	(58 - 70)	27 4
260 5	55 ('27 5)	(233 - 288)	165 8
288	14 ('7)	281 295)	12 1
288	10 ('5)	(283 293)	34 1
295	14 ('7)	(288 302)	48 3
295	10 ('5)	(290 - 300)	69 7

Figure 4.8-C-4. Table of augmentation data.

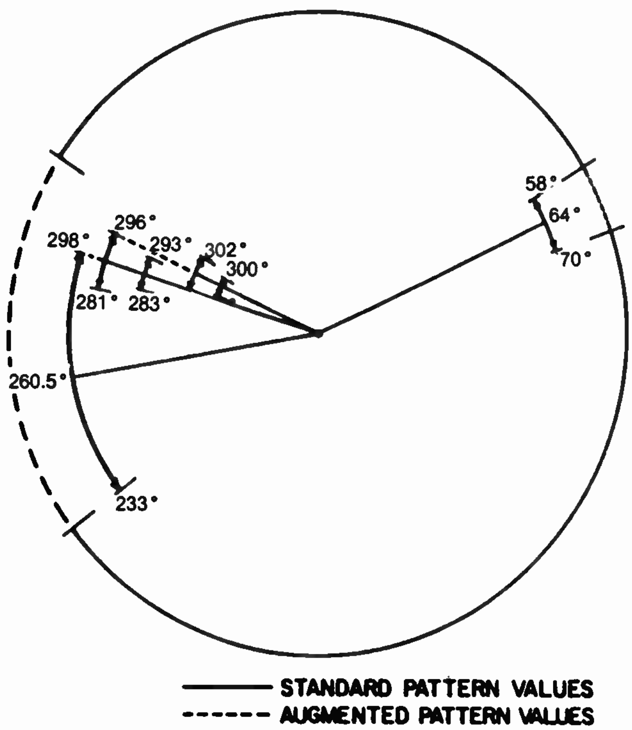


Figure 4.8-C-5. Augmented pattern flow chart showing overlapping spans.

4.9
AM ANTENNA COUPLING AND PHASING SYSTEMS

EDWARD EDISON, P.E. & GERHARD J. STRAUB, P.E.
HAMMETT & EDISON, INC., SAN FRANCISCO, CALIFORNIA

INTRODUCTION

AM broadcast antenna coupling and phasing systems consist principally of passive networks of inductors, capacitors, transmission lines, and auxiliary components such as lightning gaps, meters, jacks and relays. This equipment can range from simple (such as a nondirectional station having the transmitter building located adjacent to the tower base) to complex (such as a multi-tower directional array requiring different patterns day and night). To understand AM antenna systems and their environment, we must consider the function of such systems and their performance objectives, as well as the characteristics of basic networks, power dividers, transmission lines, sampling systems, detuning systems, and transmitter load optimization.

It is not the intent of this chapter to make directional antenna experts out of its readers. However, in providing a broad perspective of the design of such systems, we hope to encourage a rational approach to the everyday problems of directional antenna operation and maintenance. The design and adjustment of the more complex directional arrays require considerable knowledge and experience that is beyond the training of almost all chief engineers and contract engineers. When unusual problems arise that do not yield to reasonable efforts at repair and adjustment, it is time to call in an expert; preferably the consulting engineer or firm that designed or previously adjusted the system.

THE FUNCTION OF A DIRECTIONAL ANTENNA PHASING SYSTEM

The function of a directional antenna phasing system is to distribute current to each tower with controllable phase and amplitude so as to generate the desired directional radiation pattern. This function is accomplished typically by means of the following:

1. A network at each tower matches each tower load to its transmission line (see Figure 4.9-1). This network is typically in a weatherproof box, termed an antenna coupling unit (ACU) or antenna tuning unit (ATU), or within a small building called a *doghouse.*
2. The remaining equipment, usually housed in one or more indoor cabinets, is termed a *phasor.*

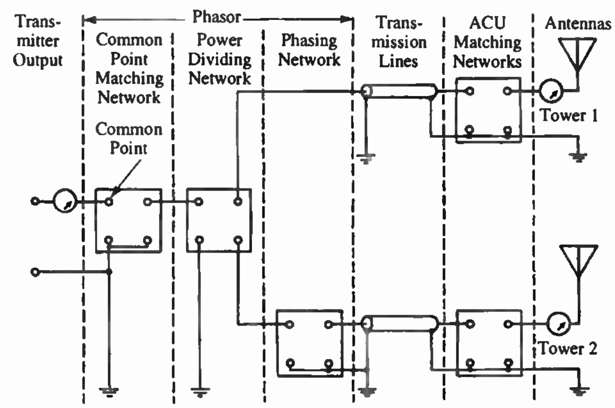


Figure 4.9-1. Block diagram of a directional antenna.

a) Within the phasor cabinet (usually located in the transmitter building), networks control the phase of the current into each transmission line.
b) Power divider circuits control the current amplitude in each transmission line.
c) A common point matching network adjusts the phasor input impedance to a desirable resistive value (typically 50 Ω) without disturbing the phase or amplitude of the tower currents.

Figure 4.9-1 shows a system with only two towers—each additional tower requires its own antenna coupling unit, transmission line, and phasing network, as well as additional components in the power divider.

PHASOR PERFORMANCE OBJECTIVES

The essential performance objective is to have both the phasor input impedance and the directional antenna radiation pattern remain essentially constant at all frequencies within ± 10 kHz of the assigned channel. The performance may be disappointing if the design process proceeds by a piecemeal approach, where subsystems are independently designed for each of the functions illustrated in Figure 4.9-1 and are then connected together. It is better to use a systems approach that considers the entire collection of components between the transmitter and the towers as a whole rather than by the *building block* approach implicit in Figure 4.9-1. No one subsystem in an array can be treated in isolation. While it is a simple matter to design a series of subsystems to provide the desired phase and amplitude parameters, these subsystems must properly complement each other if optimum performance is to be achieved.

Phasor Input Impedance

An ideal phasor would present a load to the transmitter that is the same pure resistance at all sideband frequencies. However, phasors consist of coils and capacitors whose reactance varies with frequency; therefore, this ideal cannot be achieved. The term *impedance bandwidth* describes the degree of constancy of the phasor input impedance across the entire range of sideband frequencies. A useful figure of merit to express the impedance bandwidth as a single number is the worst case voltage standing wave ratio (VSWR) existing at either of the two 10 kHz sideband frequencies when the system is perfectly matched at the carrier frequency. This approach yields VSWR numbers that are related to antenna performance just as television and FM transmitting antenna performance is described by VSWR limits at various sideband frequencies. The sideband VSWR can be determined by analysis of common point impedance measurements. The resistance and reactance at each sideband frequency of interest are expressed as a percentage of the measured carrier frequency resistance, usually 50 Ω, and are then plotted on a Smith Chart. A Smith Chart depicts the relationship between the load impedances at sideband frequencies and the ideal matched load. The distance from the center of the chart for each sideband frequency is a measure of the VSWR at that frequency. Figure 4.9-2 shows a Smith Chart showing a VSWR circle of 2.0. All impedances, when normalized to the chart center, resulting in a VSWR of 2.0 or less will plot within this circle of radius 2.0.

Bandwidth

Bandwidth requirements for each channel change greatly across the broadcast band. At 1600 kHz, a total

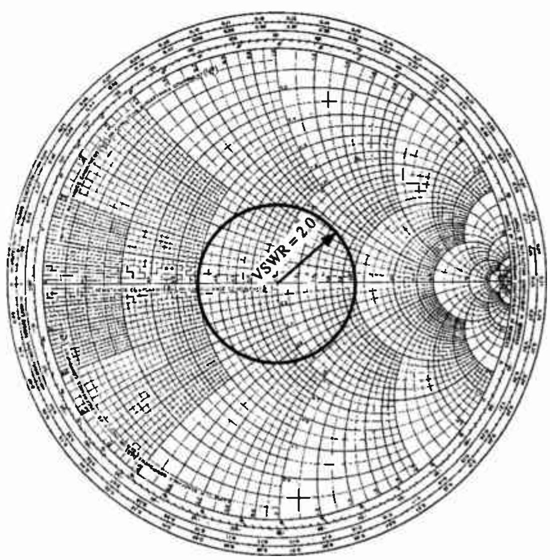


Figure 4.9-2. Smith Chart with VSWR circle.

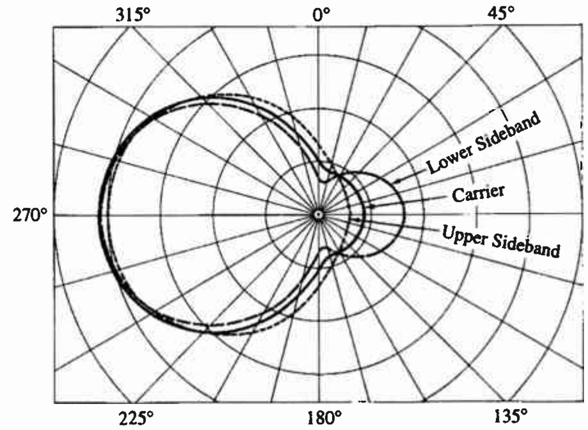


Figure 4.9-3. Example of poor pattern bandwidth.

bandwidth of 20kHz corresponds to about 1% of the center frequency whereas, at 540 kHz, 20 kHz corresponds to nearly 4% of the center frequency. Comparatively, an AM station at the lower frequencies has a more stringent bandwidth requirement than is needed for acceptable visual transmitter performance on VHF TV channels 7 through 13 or on any of the UHF channels.

The common point VSWR at 10 kHz above and below carrier is often better than 1.1:1 in good phasor designs. Poor designs can be worse than 3:1. Poor phasor designs manifest themselves by poor audio performance, which may result in limited frequency respons.

Radiation Pattern

An ideal phasor would produce an antenna radiation pattern that remains unchanged across the channel. Variations in the phase and ratio of tower currents at the sideband frequencies can result in changes in the location and depth of the intended pattern minimums. The frequencies for which the pattern remains useful describe the *pattern bandwidth.* Pattern bandwidth cannot be quantified by simple numbers, but the concept is useful in comparing alternative designs.

The example of poor pattern bandwidth shown in Figure 4.9-3 is for the three tower daytime pattern of an actual station at its carrier frequency and at the 10 kHz sideband frequencies. The sideband patterns were calculated from the actual measured ratios and phases of tower currents at both sideband frequencies. This directional antenna system did not adequately protect an adjacent channel station even though conventional field strength measurements indicated proper operation. A spectrum analyzer displaying the received signal confirmed the extreme sideband asymmetry in critical directions. In deep pattern minima, poor pattern bandwidth can result in severe distortion of the audio signal. Essentially, the carrier has been removed and what is left is effectively a distorted double sideband, suppressed carrier signal that will have an audio effect at the receiver similar to overmodulation.

BASIC NETWORKS

The networks used for matching impedances and for shifting phase are typically "T" or "L" configurations. "Pi" networks having shunt elements at the input and output and a central series element can be made electrically identical to any T network but T networks are easier to adjust.

The networks at the left in Figure 4.9-4 offer large phase shifts, while the configurations on the right offer small phase shifts. Intermediate between these two conditions are L networks which can be considered as T networks with one zero reactance arm. L networks do not permit independent adjustment of impedance match and phase shift; however, the resulting phase delay or advance can be calculated easily, with the result often being a value that is compatible with the overall phasor design.

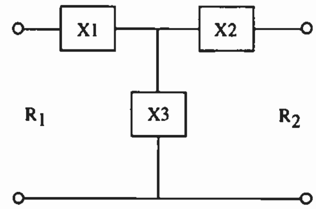
The formulas for these networks are all based on matching into resistive loads. The loads presented by antenna towers usually have a reactive component; therefore, the output arm of the matching network is modified from the calculated value to cancel the tower reactance. When this modification is made, the output arm will occasionally assume the opposite reactance sign, as shown in the parentheses in Figure 4.9-4.

T Networks

The reactances of the input, output and shunt arms of a T network can be calculated quickly, once the desired input and output resistances and the phase shift (β) are specified, by using the formulas shown in Figure 4.9-5. The formulas are equally applicable for leading ($+\beta$) and lagging ($-\beta$) networks.

L Networks

Conventional L networks can provide a match between any two resistance values. The formula for such networks is shown in Figure 4.9-6, which presumes a non-reactive load. If an L network is used to match a tower to a transmission line, the reactive component of the tower impedance must be considered. If the tower resistance is lower than the line impedance, the series arm of the L network is connected to the tower.



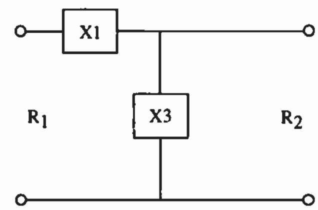
$$X_3 = \frac{\sqrt{R_1 R_2}}{\sin\beta}$$

$$X_2 = \frac{R_2}{\tan\beta} - X_3$$

$$X_1 = \frac{R_1}{\tan\beta} - X_3$$

R_1 = Input resistance
R_2 = Output (load) resistance
β = Phase shift (output - input)

Figure 4.9-5. "T" network formulas.



$$X_1 = \pm\sqrt{R_1 R_2 - R_1^2}$$

$$X_3 = \pm\frac{R_1 \; R_2}{X_1}$$

$$\beta = \cos^{-1}\sqrt{\frac{R_1}{R_2}}$$

R_2 must be greater than R_1
X_1 is inductive (+J) for lagging network
X_1 is capacitive (-J) for leading network
X_3 must have sign opposite to X_1

Figure 4.9-6. "L" network formulas.

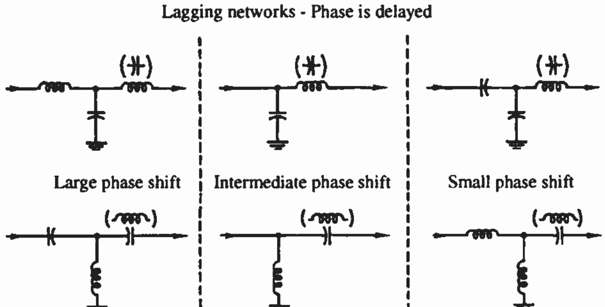


Figure 4.9-4. Conventional "T" and "L" networks.

This arm can then be modified from the value calculated by the formula so as to also cancel the tower reactance.

An L network can also be used to match a tower having a resistance higher than the transmission line impedance, but in this case the formula in Figure 4.9-6 does not apply unless the tower impedance contains no reactance. In the usual case, where the tower impedance is reactive, an L network cannot be so easily calculated. The shunt arm (which is in parallel with the reactive tower load) must be one which makes the resistive component of the parallel combination equal to the transmission line impedance. Then the resulting reactance of the parallel combination is canceled by the series input arm of the L network. Such networks are easy to adjust; the shunt arm of the L network is adjusted to make the input resistance match the transmission line characteristic impedance, and the series arm is then adjusted to cancel the resulting reactive component. The two adjustments are substantially independent of each other. If the tower is one element of a directional array, an operating impedance bridge (hot bridge) at the network input must be used and the phase and ratio parameters must be approximately correct in order for the tower operating impedance to be that which will exist when the directional antenna array is operating normally.

When the tower resistance is higher than the line impedance, there are two different values of shunt arm reactance (one capacitive and one inductive) that can satisfy the conditions necessary for a match with an L network. If the shunt arm is of the same sign as the tower reactance, the required series arm will be of opposite sign and the phase shift will be small. If the shunt arm is opposite in sign to the tower reactance, the required reactance in the series arm will have the same sign as the tower reactance and the phase shift will be large.

An alternative way to design such L networks (rather than by calculating the shunt arm reactance necessary to produce the desired resistive component for the parallel combination) is to think of them as T networks in which the output arm is represented by the reactive component of the tower load. As such, these networks can be described as *Phantom T Networks.*

Phantom T Networks

A phantom T network utilizes the tower reactance itself as the output arm (X_2) of the network. It is electrically equivalent to a conventional T network but requires one less circuit element. Because it has only two adjustable elements, the phase shift of a phantom T cannot be independently selected but two choices are possible, corresponding to T networks with small and large phase shifts. When the input resistance and reactance are properly adjusted, the resultant phase shift is then defined. Phantom T networks can be calculated quickly by the T network formulas while using cut-and-try variations of the phase shift until X_2 exactly equals the tower reactance. Phantom T networks with small phase shifts exhibit excellent bandwidth. Those with large phase shifts are somewhat poorer.

POWER DIVIDERS

In many phasors, separate components are used to accomplish the power division function; but the same networks that control phase shift can also control the division of power, if desired. It is important to remember, when comparing different towers in an array, that the power delivered to a tower is not in proportion to the base current; a low base resistance tower may consume very little power (I^2R) even if the current is relatively high. The actual power division is defined by the required base current ratios and the base resistances. An example of power division calculations is shown in Figure 4.9-7.

Given: Total power is 1000 watts with base current ratios and resistances as shown. I = Current in Tower 3.

Tower No.	Base Current Ratio	Base Resistance	Power
1	0.20	25 ohms	$(0.20\ I)^2 \times (25) = 1.0\ I^2$
2	0.70	130	$(0.70\ I)^2 \times (130) = 63.7\ I^2$
3	1.00	40	$(1.00\ I)^2 \times (40) = 40.0\ I^2$
4	0.50	10	$(0.50\ I)^2 \times (10) = 2.5\ I^2$
			$107.2\ I^2 = 1000$ watts; $I = 3.054$ amps

Tower No.	Base Current	Base Power
1	0.20 x 3.054 = 0.611 amps	$(0.611)^2 \times (25) =$ 9.3 watts
2	0.70 x 3.054 = 2.138	$(2.138)^2 \times (130) =$ 594.2
3	1.00 x 3.054 = 3.054	$(3.054)^2 \times (40) =$ 373.1
4	0.50 x 3.054 = 1.527	$(1.527)^2 \times (10) =$ 23.4
		1000.0 watts

Figure 4.9-7. Example of power division calculations.

Tank-Type Divider

In a tank type power divider, as shown in Figure 4.9-8, roller coils L2 and L3 (termed *jeep* coils in this configuration) adjust the power; the combination of L1, L2, and L3 is tuned close to resonance by capacitors C1 and C2. The input tap on L1 adjusts the common point resistance; L4 adjusts the common point reactance. This system is no longer popular because it is difficult to adjust and because bandwidth is limited by the excessive stored energy.

Shunt-Type Divider

A common type of divider is the *shunt* design shown in Figure 4.9-9, which uses several adjustable coils in parallel, each of which controls the current to an individual tower. The reactance of the L1, L2, L3 combination is then tuned to resonance with C1. Component values are selected so that the impedance at point A is on the order of 400 Ω. A conventional T network then matches this load to the desired common point impedance. This design is easy to adjust. L2 and L3 may be tapped or variable inductors, however, the rollers on the variable coils may tend to bind and develop intermittent contacts after several years of use.

T Network Power Division

A phasor can be designed to exploit the input impedance of conventional T networks to achieve control of both power division and phase, as shown in Figure 4.9-10. In such a system, each T network presents a load at its input which is appropriate to divert the required power to its corresponding tower.

Once such a phasor is properly adjusted, the small changes in phase and ratio that may later be required to maintain an array within tolerances are easily accomplished with very small adjustments in only the shunt and input or output arms of each individual tower network in the phasor. Although the phase and ratio adjustments interact (as with all phasors, due to the

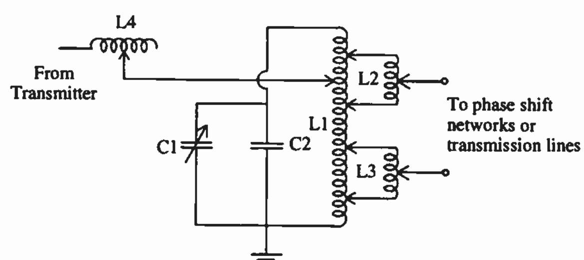


Figure 4.9-8. Typical tank-type power divider.

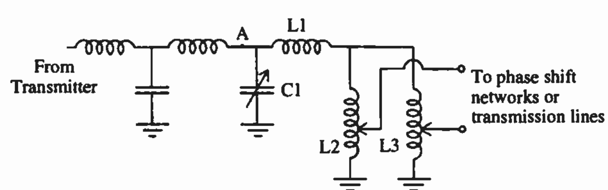


Figure 4.9-9. Typical shunt power divider.

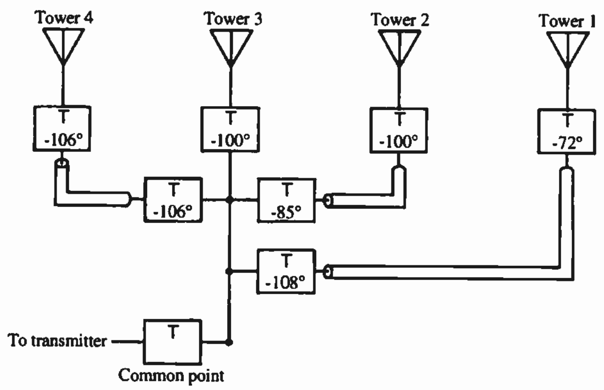


Figure 4.9-10. Typical phasor with "T" net power division.

mutual coupling between towers), the shunt arms tend mostly to control power division, whereas the series input arms tend mostly to control phase. By noting which parameters are farthest out of tolerance, it is a simple matter to make small adjustments in the networks, observe the results of each change, and expeditiously restore such a phasor to the desired parameters.

Simplified Power Division

For two tower directional systems, some very simple configurations can permit adjustment of both power division and phase shift with very few circuit elements. One such simplified system is a *back-to-back* phasor, shown in Figure 4.9-11.

Components "C" and "L" are connected back-to-back at the output of the common point matching network. In this example (which assumes matched transmission lines with resistive inputs), reducing L to zero would yield a current into transmission Line 2 that is in phase with the applied voltage, E_0. As L is increased, the current into Line 2 diminishes and lags the applied voltage. As L is varied from zero to infinity, the locus of all possible Line 2 currents is the semi-circle on the lower half of the diagram. Similarly, as C is reduced, the amplitude of the current in Line 1 is reduced and its phase is advanced relative to the

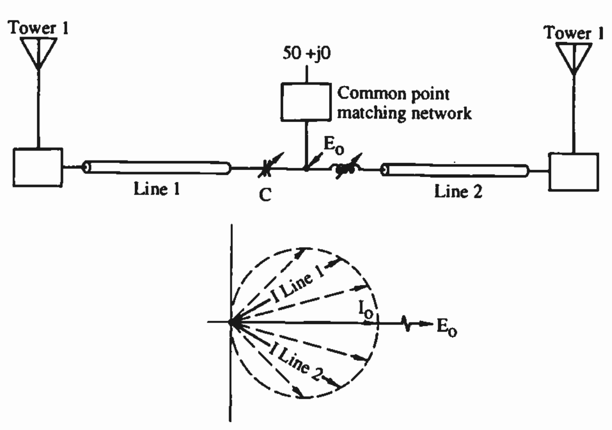


Figure 4.9-11. Example of back-to-back phasor.

applied voltage. The locus of all the possible currents into Line 1, as C is varied from infinity to zero, is the upper semi-circle of the vector diagram. The ranges of adjustment of C and L need only be sufficient to reach the desired phase and ratio parameters, not the entire semi-circular gamut.

Independence of adjustment of phase and ratio is completely lacking in a back-to-back design, yet the phasor is easy to adjust. If one wants to change the ratio, an arbitrary change is first made in either component, phase is restored to the initial value by adjustment of the opposite component, and then the direction of ratio change is noted. It becomes a very simple cut-and-dry process, not unlike that required to maximize transmitter efficiency by adjusting transmitter tuning and loading.

It is not always necessary to employ a coil and a capacitor in a back-to-back phasor. Often, the phase shift in the transmission lines and ACU networks can be tailored to provide the desired parameters with a series capacitor on the input to each line, rather than one capacitor and one coil. This arrangement permits a two-tower phasor in which both power division and phase adjustment are accomplished by only a pair of variable vacuum capacitors offering excellent stability and resetability. This arrangement also provides excellent impedance bandwidth.

TRANSMISSION LINES

Many of the requirements for AM broadcast transmission lines differ from those used in FM and TV installations; the similarities and differences are discussed in the following paragraphs.

Line Losses

The losses in transmission lines at AM broadcast frequencies are largely in the copper, not in the dielectric. For this reason, and for mechanical convenience, semi-flexible foam filled lines are the most popular. Air dielectric lines are needed only when power levels exceed the ratings of available foam filled lines.

Jacketing

All lines should include a black polyethylene outer jacket. Unjacketed lines can result in inadvertent poor contacts to other metallic conductors, which can be a source of minute arcs that can cause spurious emissions. Jacketed lines are particularly suitable for direct burial. Burying the lines greatly reduces the daily and seasonal temperature extremes they encounter, yielding greater stability relative to above ground installations.

Although jacketed lines have an indefinitely long life when buried, they are vulnerable to attack by rodents, that can gnaw through the vinyl jacket and have been known to consume portions of the copper outer conductor, as well. The possible need to replace buried lines in the future should be considered at the time of initial construction. Protecting the lines in PVC conduits and/or installing spare lines are desirable options. At the minimum, PVC sweep elbows should be in place before concrete floors or foundations are poured for buried transmission lines and sampling lines that exit the transmitter building or doghouses.

Characteristic Impedance

A characteristic impedance of 50 Ω is the industry standard for radio frequency transmission lines. Such lines can be expected to be more easily available than other impedances in future years, should a replacement line be needed. The velocity of propagation in transmission lines is less than the velocity in air. Therefore, the electrical length of any line is somewhat greater than its physical length. This increase must be considered in any phasor design. Typical values of transmission line velocity constants are listed below:

Obsolete solid dielectric flexible lines	68%
Original foam filled lines	79%
Low density foam (LDF) lines	88%
Semi-flexible air dielectric lines	90–93%
Modern rigid copper lines	99.8%

Phase Stabilized Line

In its first year, a new directional array will usually experience small drifts in indicated phase and ratio parameters that will not be repeated seasonally thereafter. This initial drift is caused by the minor mechanical stresses remaining in the feed lines and sampling lines following manufacture and installation; these stresses are slowly relieved with temperature cycling. The result is a small initial change in characteristics following installation that will not be repeated. Phase stabilized lines, which will reduce the initial drift, are available. Such lines have been temperature cycled in an oven to relieve residual mechanical stresses. Within about a year following installation, no observable differences in stability exist between regular lines and those that were initially phase stabilized.

Transmission Line Fittings

All modern transmission lines can be equipped with EIA standard flanged fittings to mate with connecting equipment. Although the impedance continuity provided by such fittings is essential at FM and TV frequencies, such fittings are unnecessary for foam line antenna feeds at AM broadcast frequencies. However, adequate electrical bonds are essential between the transmission line conductors and the adjoining equipment. The dielectric in foam filled lines is a better insulator than air. However, catalogs show such lines as having the same voltage breakdown rating as air dielectric lines because of the limited air gap within the end fittings. For AM broadcast use, the power rating of foam lines can be increased by avoiding the use of end fittings and stripping back the outer conductor (while leaving the foam) to increase the air gap on both ends of each line.

Line Mismatch

At AM broadcast frequencies, the effects of a transmission line mismatch are quite different from the electrically long transmission lines used for FM and television antennas, where mismatches can cause reflections back into a transmitter and result in crosstalk or ghosts. At AM broadcast frequencies, transmission lines are rarely longer than a wavelength. A mismatched transmission line yields an input impedance that is dependent on both the load impedance and the length of the line. Thus a transmission line can act as a simple impedance transformer and can be designed to exhibit a desired input impedance through proper choice of load impedance. This characteristic is occasionally exploited in designing phasing systems.

Perfect transmission line matches are not critical to overall system bandwidth. Although most phasing and coupling systems are designed with the intent that the transmission lines to the towers be matched, the consequences of substantially mismatched lines are usually trivial. Improving the match on any transmission line between the phasor and tower may improve or may impair the bandwidth. Because the lines are electrically short, their behavior is more akin to that of a simple network than to the transmission lines in FM and television systems, which are often hundreds or thousands of wavelengths long. As a general rule, if lines to towers that handle 25% or more of the transmitter power are matched to a VSWR of 1.5:1 or better (at the carrier frequency) and any remaining lines are matched within 2:1, further efforts at improving matches may give only trivial reductions in line losses and trivial changes (for better or worse) in the bandwidth.

Even though a line may be perfectly matched at the carrier frequency, it will never be perfectly matched at the sideband frequencies because no antenna load is absolutely constant over the desired bandwidth of ±10 kHz. The effects of the less than perfect sideband matches can be minimized by proper proportioning of the phase shifts between the phasor networks and the ACU networks. It is important to avoid confusion between transmission line VSWR at the carrier frequency and common point VSWR at sideband frequencies.

Voltage and Power Ratings

With 100% modulation, the voltage on a matched lossless transmission line varies from zero on a modulation trough to twice the unmodulated value on modulation crests. This is quite unlike the situation in FM where every radio frequency cycle has the same amplitude. In FM and TV systems, the dielectric losses (because of the higher frequencies) cause the power rating to be limited by the heat generated within the dielectric, as well as by the copper losses; power ratings are accordingly based on the ability of the transmission line to dissipate the heat resulting from the losses. At AM frequencies the dielectric losses are minor while the instantaneous voltages during modulation peaks can be very high. As a result, the power handling ability of AM transmission lines is usually limited by the voltage breakdown rating of the line. With 125% positive modulation, the maximum instantaneous peak voltage becomes more than three times the unmodulated RMS voltage and the instantaneous peak power becomes more than 10 times the unmodulated power. An additional allowance must also be made to accommodate the increased voltage maximum along each line due to standing waves caused by any less than perfect transmission line match.

Under transient conditions (such as can be caused by lightning strikes or sudden component failures) the VSWR overload protection designed within all modern transmitters may not trip and the power distribution among the towers may be drastically altered in unpredictable ways. This can lead to voltage breakdowns within buried transmission lines which are difficult to locate and repair. For this reason it is good engineering practice to employ equal size transmission lines to all towers that are capable of handling the full modulated output of the transmitter even though the power delivered to particular towers may be only a few percent of the total power during normal operation. This design factor will also be greatly appreciated in the event of an emergency requiring non-directional operation on one of the low power towers.

ANTENNA MONITOR AND SAMPLING SYSTEM

A directional antenna sampling system consists of current sampling transformers at the base of or sampling loops on each tower to provide a sample of the tower current, transmission lines which return the samples, and an antenna monitor which measures the amplitude and phase of each sample relative to that of the reference tower.

Sampling Loops

A sampling loop consists of a rigid single turn coil permanently attached to each tower. Because sampling system stability is essential, the loops are typically made of galvanized or stainless angle iron or rigid copper water pipe. The loops must be at least 10 ft. above ground and may be insulated from the tower and kept at ground potential. However, it is more common practice to use loops at tower potential and return the sample to ground potential through an isolation coil (*isocoil*) formed by coiling up the sampling coax so as to form a high impedance across the tower base insulator while not disturbing the current sample carried within the coax.

Sampling Transformers

An alternative method of obtaining tower current samples is to use shielded current sampling transformers in the feeds to each tower. These have much to recommend them if the tower heights do not exceed approximately 130°. Although a transformer will sample both the current going up the tower and a lesser component that flows to ground through the base insulator capacitance, the sampling error is usually inconsequential. The advantages to using sampling transformers are stability, a sampling device that is protected inside the ACU instead of being exposed to the weather, and elimination of the isocoils that are otherwise usually required at each tower.

Stability and Accuracy

Sampling system stability is vital, but accuracy is not critical because the proper phases and currents to generate a directional antenna pattern are determined by means of field strength measurements. The sampling system is required to detect changes from the original adjustments of indicated phase and ratio. The accuracy of a sampling system in measuring the absolute amplitude of radiation from each tower is inherently poor.

Even though the sampling loops may be carefully adjusted so that each has exactly the same area in its single turn coil and all are mounted at the same height, the samples will not necessarily be a measure of the relative radiation from each tower. The mutual impedances between towers can seriously distort the sinusoidal current distribution which would otherwise exist throughout the height of each tower. Thus the indicated antenna monitor ratios are not an absolute measure of the relative radiation from each tower. The monitor ratios may also differ substantially from the base current ratios, particularly when the loops are a considerable distance above ground level or towers of unequal height are employed. The indications of phase from a good monitoring system are considerably more accurate than the ratio indications. Arrays with unequal height towers usually require sampling loops that are all located approximately a quarter wavelength below each tower top if the indications of phase and ratio are to be the best approximation of the actual radiation from each tower.

Recently, there has been some activity directed at simplifying the proof-of-performance of directional arrays by eliminating or at least reducing the requirement for field strength measurements. With the computer power available today, it is possible to calculate relatively accurately the current distribution on a given tower in an array and to determine in advance what monitor values are needed to produce the required pattern. In order to do this, the absolute accuracy of the sampling system becomes an important concern. As a result, it would be necessary to characterize the sampling system prior to array adjustment so that the target monitor values could be adjusted accordingly. Even with a perfect sampling system, however, some field strength measurements would be necessary, as the sampling system can not determine the effect of external factors located near the array.

DETUNING AND DECOUPLING SYSTEMS

The most common occasion for detuning towers arises within directional arrays themselves, wherein unused towers must be placed in a non-radiating condition. This is necessary in all arrays when nondirectional proof-of-performance field strength measurements

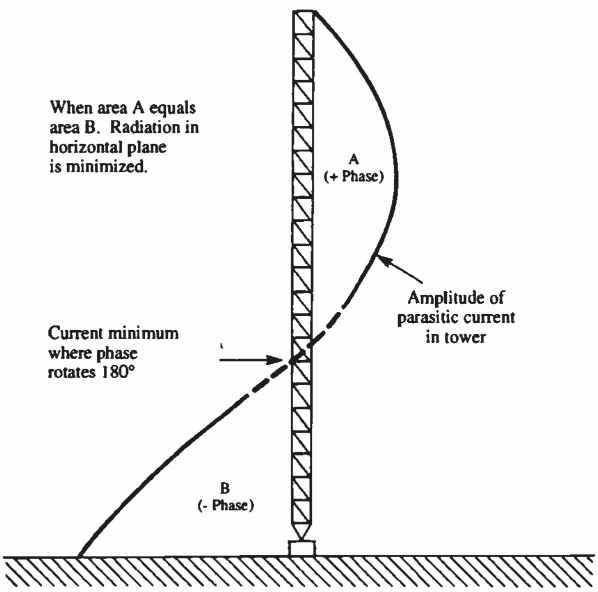


Figure 4.9-12. Current distribution on detuned tower.

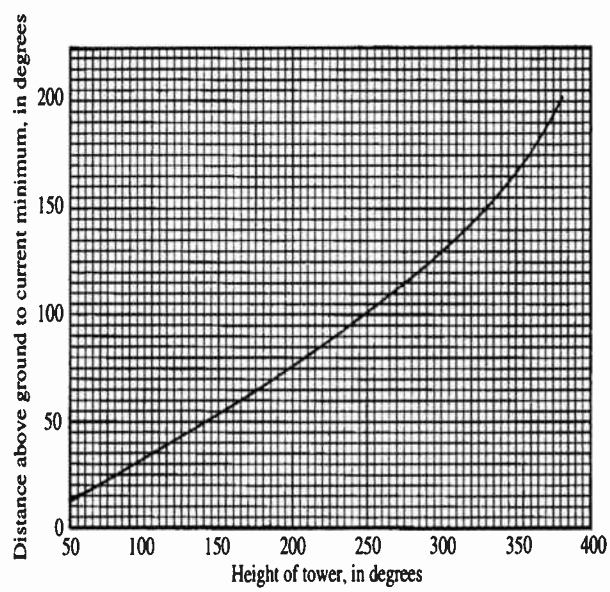


Figure 4.9-13. Approximate height of required current minimum to detune guyed towers.

must be made, where nondirectional daytime operation is permitted, or where one or more towers are not used for daytime or nighttime modes when two patterns are authorized. In more critical situations, or with taller towers, more effective detuning is required than simply disconnecting the unused towers to reduce the radiation (reradiation) caused by parasitic currents that flow in them due to the mutual coupling to adjacent driven towers. Disconnecting a tower will not eliminate the component of induced tower current flowing to ground through the capacitance of the tower base and base insulator. An inductive reactance of the proper value (often formed in part by the isocoil if sampling loops are employed) can be added across the base to resonate the base capacitance and essentially eliminate all current flowing to ground.

The optimum detuning (even for short towers, as well as for towers up to a wavelength tall) results when an inductive reactance across the base insulator produces a tower current distribution as shown in Figure 4.9-12. Because there is a phase reversal at the elevation of minimum current, it is necessary to position a current minimum on the tower so that area A equals area B. In the horizontal plane, the radiation from the portion of the tower above the minimum will then cancel that from the portion below the minimum. The proper height for the tower current minimum (typically about 1/3 the tower height) is as shown on Figure 4.9-13. The minimum can be achieved easily by climbing the tower with a field strength meter (or other current sampler and detector) and adjusting the reactance shunting the base insulator until a current null is obtained. Although the shunt reactance necessary to detune a tower is always inductive, the reactance may be so high as to not be within range of a coil of reasonable size. A vacuum variable capacitor in parallel with the coil then provides a convenient and precise way of adjusting the shunt reactance so as to detune the tower.

Tower Sectionalization

Occasionally, when it is necessary to use a tower that is too tall, sectionalization of the tower is required. For example, it may be desired to share an FM tower with an AM station. If the overall tower height were to exceed about 225° at the AM frequency, the tower would be a poor radiator, putting more AM signal up into the sky than along the ground. The preferred solution for such a tall tower is to include insulators in the tower legs (and an isocoupler in the FM feedline at the same height) so as to open circuit the tower at the height needed to make it a good AM radiator. A parallel resonant circuit (an impedance pole) across the insulators can then resonate the insulator and stray capacitances to effectively decouple the upper portion of the tower. In cases where structural considerations prevent the sectionalizing of a tower by means of insulators, a workable solution can be to use detuning *skirts* to decouple the top portion of the tower.

A tower skirt is an insulated wire cage constructed outside of a tower. The skirt is formed of wires that surround the tower and are spaced a foot or two away. All skirt wires are bonded to the tower at one end and can be connected in parallel at the other end so as to be tuned with a single capacitor. Electrically, the skirt can be considered as forming a shorted quarter wave coaxial transmission line which is slightly foreshortened and loaded with a variable capacitor at its open end for easy adjustment. The tower forms the center conductor of the coax and the skirt wires approximate an outer conductor. Adjustment of the skirt tuning places a high impedance at the capacitor end of the

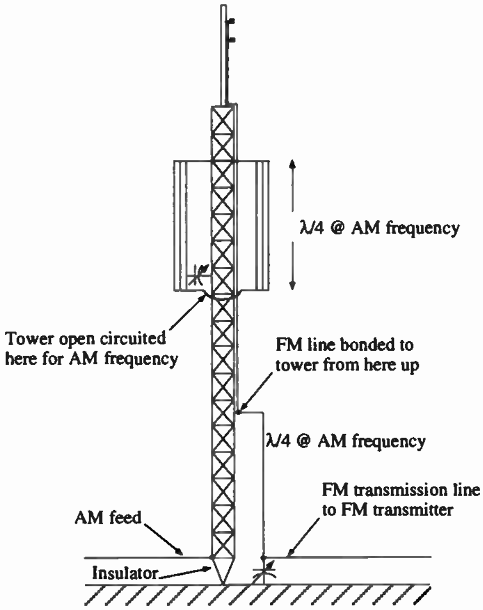


Figure 4.9-14. AM/FM tower sectionalized with skirt wires and with decoupled FM transmission line.

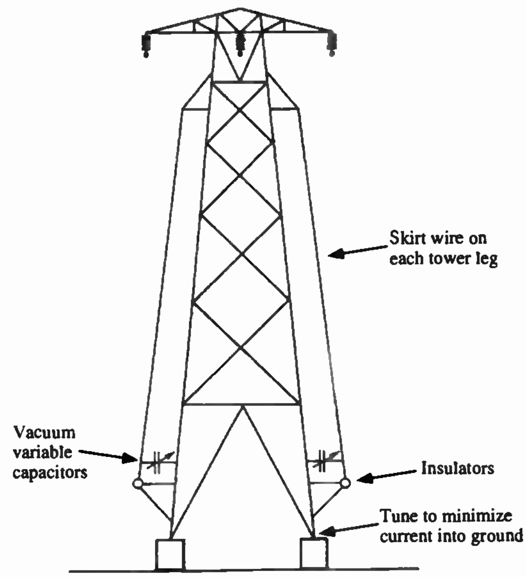


Figure 4.9-15. Skirts used to detune power-line towers.

skirt with a result that the tower is effectively open-circuited at that elevation.

The effectiveness of such a tower skirt is determined by the characteristic impedance and losses in the transmission line section formed by the tower and the skirt. The most effective skirt will necessarily have a high Q, which is realized by using reasonably heavy skirt wires spaced as far as mechanically feasible away from the tower. For guyed triangular towers of uniform cross section, combinations of three or six skirt wires are common. Inasmuch as the most effective skirt will necessarily have a high Q, it will best decouple the portion of the tower above its open end only at the frequency to which it is tuned. As a result, the tower base may exhibit a steep impedance versus frequency characteristic which can impair bandwidth.

Another way of visualizing the operation of a skirt is to think of the tower and skirt wires as forming an inductive loop which is tuned to resonance so that the current flowing up the tower is effectively canceled by a substantially equal and opposite current flowing down the skirt wires. Figure 4.9-14 shows an example of the use of skirt wires to sectionalize a tower and to decouple a transmission line for the FM station located at the top.

Detuning Power Lines

The detuning techniques previously described can be applied to steel power line towers or to other tall metallic structures in order to reduce the radio frequency currents in them, currents that would otherwise cause reradiation and distort the desired directional antenna pattern of a nearby radio station.

Figure 4.9-15 shows such an arrangement with each skirt wire on each tower leg separately tuned. Again, the effectiveness of the skirt depends on the length of the skirt section and the number, diameter and spacing of the skirt wires.

Filter Systems

If it is necessary to control the current distribution on an AM tower at a different frequency from the operating frequency in order to avoid disturbing the pattern of a nearby directional station, a filter system is needed to permit control of the impedance shunting the tower base at the other station's frequency. The control should be independent of routine phasor adjustments. Such an arrangement is shown in Figure 4.9-16 for each tower of a station operating on 630 kHz. The 630 kHz station must control tower current distribution at 950 kHz in order to protect the directional antenna operation of the nearby station on that frequency.

For filters in series with antenna feeds (such as the C3/L7, C4/L8 combination), it is desirable to provide an impedance zero at the pass frequency as well as the impedance pole at the reject frequency. Even when

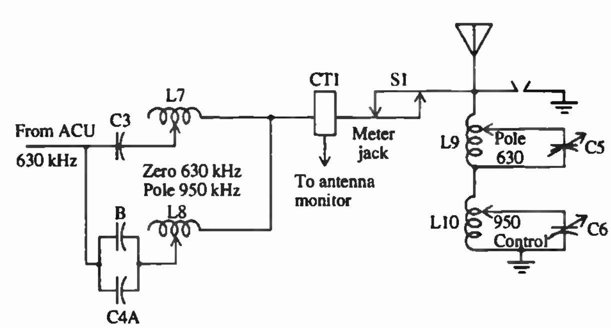


Figure 4.9-16. Independent control of tower current distribution at a second frequency.

the pass and reject frequencies are widely separated, failure to provide the impedance zero with a reasonably low L/C ratio can result in excessive losses.

Intermodulation

Filters to control spurious emissions resulting from intermodulation products generated within the final stage of a transmitter can have the same configuration as filters used to control tower current distribution for a nearby station. Before such filters are designed, a knowledge of the frequencies that generate the intermod is essential. The second, third, and higher harmonic frequencies are strong components in a transmitter output stage and can generate intermod products there even though these harmonics are suppressed to acceptably low levels when measured at the transmitter output. In the general case, an unwanted incoming signal from some other station can be coupled into the final amplifier through the antenna system to mix with the fundamental or any harmonic to produce a spurious output frequency.

Effective filters can take the form of either traps to prevent an unwanted signal from getting into the transmitter or traps to prevent the resulting spurious emission from getting out (or both). Filters can be series resonant shunt elements that short the unwanted incoming signal (or the spurious emission) to ground, or they can be parallel resonant elements in series with the transmitter output to present a high impedance to either the unwanted incoming signal or the spurious emission.

One clever and expeditious arrangement, when a single spurious frequency must be eliminated, may be to modify the shunt arm of the common-point matching network to consist of a series-resonant L/C circuit that simultaneously has a very low impedance at the undesired frequency and the proper reactance at the station carrier frequency.

Many modern transmitters have exceedingly broadband output coupling networks which, together with improved antenna bandwidths, permit unwanted signals from other radio stations to get back into the final stage of the transmitter to produce spurious intermod products. Signals thus entering the transmitter need not be particularly strong. Many cases are known where the offending input that caused an objectionable spur originated from a radio station 20, 30, or even 40 miles away. Be aware that the intermodulation products may fall in the marine and beacon bands below the AM broadcast band.

Broadbanding

The term *broadbanding* can be quite confusing. Frequently it is applied to any circuit modifications that are intended to improve the impedance bandwidth of an antenna system. A *broadbander* is a circuit that improves the impedance bandwidth, thus reducing the VSWR at the sidebands. A *broadbander* typically consists of the series combination of a coil and a capacitor which is inserted in series with the load to be improved (usually immediately before or after the common point) and is adjusted to series resonance at the carrier frequency. Such a circuit is capacitive below the carrier frequency and inductive above the carrier frequency. Before a broadbander can function as intended, the load it sees must be inductive at the lower sideband frequencies and capacitive at the upper sideband frequencies. This condition is achieved by altering the phase shift in the common point matching network (or inserting an additional phase shifting network) so that the Smith Chart plot is *horns up*, as is shown in Curve A, Figure 4.9-17. Then the uncorrected load reactance (which is inductive for the lower sideband) can be canceled by the capacitive reactance of the broadbander and the capacitive reactance of the upper sideband load can be canceled by the inductive reactance of the broadbander. The result is that the load reactance can be effectively canceled at any pair of sideband frequencies as determined by L/C ratio of the broadbander. For such a *broadbander* to function as intended, it is essential that it be effectively shielded so that there is no stray coupling between it and the other components in the phasor. When properly designed and adjusted, it can *wrap up* the common point impedance, as shown in Curve B, Figure 4.9-17.

Such a broadbanding circuit should not be viewed as a cure-all, but it can provide a useful reduction in sideband VSWR once all other reasonable efforts at improvement have been applied.

TRANSMITTER LOAD OPTIMIZATION

A transmitter can deliver equal power to both sidebands when the upper and lower sideband load resistances are equal and the upper and lower sideband load reactances are equal. If the transmitter load VSWR at ±10 kHz from carrier is rather good (perhaps 1.2:1 or better), further adjustments to improve load symmetry may not be warranted. However, if the ±10 kHz sideband loads are rather poor (e.g., 1.5:1 or greater) sideband load symmetry can become an important matter. Symmetry is achieved by adjusting the phase shift in the common point matching network so that the load resistances at the ±10 kHz sideband frequencies are essentially equal and the reactances are equal but opposite in sign as measured at the anode (or collectors) of the transmitter output stage.

Because different manufacturers use different output circuitry with different phase shifts, and because some have not published information as to optimum load orientation at the transmitter output terminal, the necessary measurements usually must be made on the load seen by the final amplifier stage anode or collector. Figure 4.9-17 shows the Smith Chart orientation for the two conditions that satisfy symmetrical loading requirements. For tube type transmitters (whose output stage functions substantially like a constant current generator), Curve C, which shows equal sideband resistances higher than at carrier, tends to emphasize the high frequency response. Curve A also shows equal sideband resistances, but lower than at carrier; this orientation tends to show reduced 10 kHz distortion.

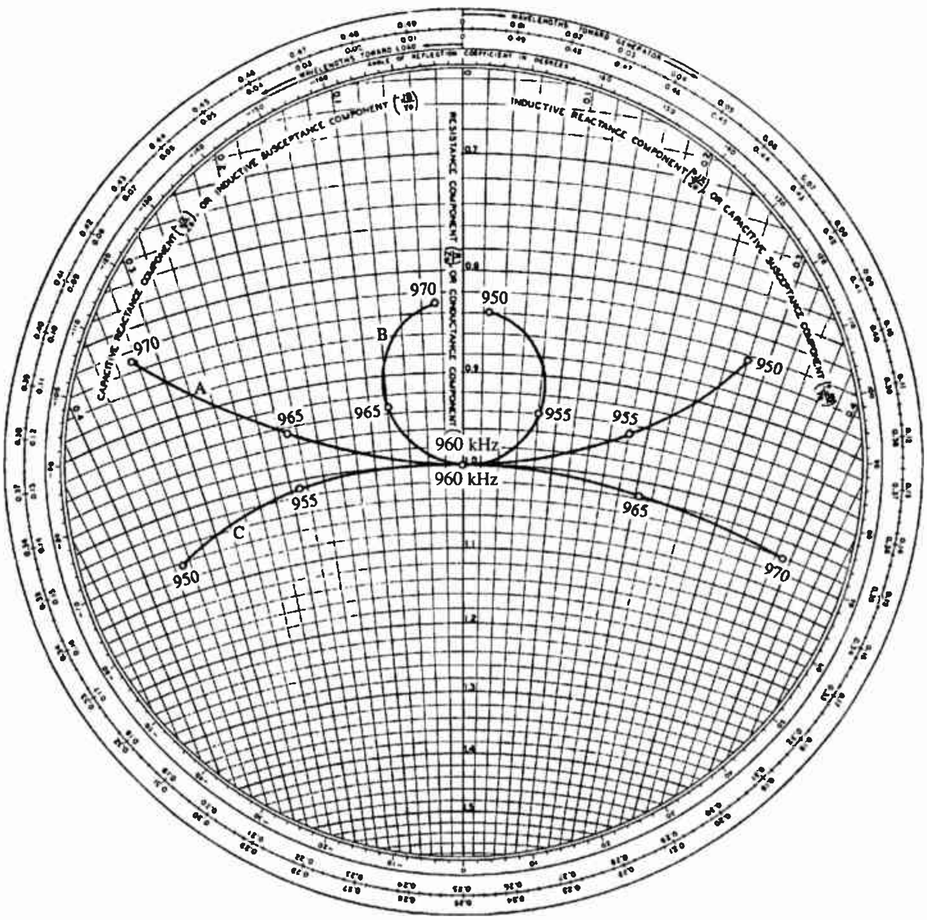


Figure 4.9-17. Proper load impedance at final amplifier plate or collector.

Either of the two conditions is preferable to intermediate conditions that yield asymmetrical load resistances at the two 10 kHz sideband frequencies.

Transmitter manufacturers are now more sensitive to the fact that all antenna systems provide less than perfect loads at the sideband frequencies and that load orientation (as displayed on a Smith Chart) can be critical to overall system performance. Most manufactures can now supply information as to optimum load orientation (referred to the transmitter output terminals) for each specific transmitter. Experience has also shown that audio intermodulation distortion varies with load orientation, so intermod should be measured and minimized.

DIPLEXERS AND MULTIPLEXERS

Two, three, or more AM stations can be combined on a single tower, provided that sufficient frequency separation exists between the channels and the electrical height of the tower is not unreasonably short or tall at any of the channels. Depending upon the transmitter characteristics, only 35 dB or less of RF isolation is typically required between the transmitters to avoid generation of spurious emissions and to permit completely independent operation. Diplexer adequacy is assured when each station:

- Has no objectionable spurious emissions.
- Can conduct an acceptable audio proof, including distortion and noise measurements, while the other station is in operation.
- Has no observable change in its indicated antenna current when the other station's transmitter is turned on and off.

The losses in diplexing and multiplexing filters can become crucial to a successful design. Since the tower impedance and required tower current at each frequency are known, the losses can be calculated with good accuracy. Capacitor losses are negligible. Coil losses can be calculated using an assumed coil Q of 300, which is a conservative approximation for typical silver-plated RF inductors in shielded enclosures. Diplexer losses can be minimized if the reactive component of base impedance can be eliminated by means of networks at the tower base. Then the shunt elements in the diplexing filters operate at lower voltage levels.

Two stations having directional antenna patterns can be diplexed on common towers provided the tower spacing and orientation permit acceptable directional patterns for both stations and the compatibility requirements previously mentioned for combining nondirectional stations are met.

Ground Systems

The current that flows into a tower does not simply disappear; it all flows back to ground through the capacitance that exists between each incremental length of the tower and the ground. (This makes the tower and ground system an integral part of the coupling and phasing system.) When this current reaches the ground, it must be collected and returned to the ground side of the ACU network to complete the circuit. The ground currents are usually strongest close to the base of the tower where the capacitance between the tower and incremental ground segments is greatest. If the ground resistance is substantial, the ground current (equal to the antenna current) will simply heat the ground and represents a power loss. To return the antenna current from the ground surrounding the tower to the ground terminal of the ACU network, a system of radial ground wires is employed. For nondirectional towers a series of early day experiments established that 120 radial ground wires uniformly distributed at 3° intervals and extending a quarter-wavelength (90 electrical degrees) from the tower base was a reasonable limit beyond which the reduced losses to be realized from incremental increases in the number or length of the radials did not warrant the incremental increases in cost required for a more extensive ground system. As a result ground systems consisting of 120 radials extending a quarter wavelength from each tower (or foreshortened and bonded to a transverse copper strap where radials from adjacent towers overlap) have become the *de facto* industry standard, even though it is recognized that in directional arrays the ground currents do not always flow along radial lines, due to the effects of adjacent towers.

Near the base of each radio tower, there are both electromagnetic and electrostatic fields. With reasonably low base impedances (such as a quarter wave tower), the electromagnetic field predominates. Near towers with a high base impedance (such as a half wave tower), the electrostatic field predominates. The radial wire ground system minimizes the I^2R losses but not the E^2/R electrostatic losses which can occur in all materials near the tower base with less-than-perfect dielectric constants. The electrostatic losses are usually inconsequential in a nondirectional operation but can become critical in a directional array. Dielectric materials in the vicinity of the tower base can include wooden fences, wooden or cement block tuning houses, and the earth that is on top of a buried ground system. The changing dielectric constant in these materials between wet and dry conditions can make enough difference in base impedance to cause substantial changes in base current ratios and phases. For these reasons the ground wires should be kept on top of the soil in the immediate vicinity of each tower base or covered with only a thin layer of gravel or paving. Beyond about 20 ft. from each tower base the fields are less intense and the remaining portions of each ground radial are usually buried 6-8 in for mechanical protection. Deeper burial is called for only if the land is to be cultivated or is so soft that farm animals or machinery can sink into the soil and break the ground wires.

Metal fences surrounding each tower base are quite superior to wooden fences, whose dielectric constant can change greatly when wet. If non-metallic doghouses are employed in critical arrays, it may be necessary to provide an electrostatic shield of expanded copper mesh or Copperweld wire mesh up the side of the dog house adjacent to the tower and across its roof. The ultimate electrostatic shield takes the form of an elevated radial ground system which typically extends to a radius of 10 to 30 ft. at a height of 6 ft. or more above ground and effectively shields everything beneath it.

Number 10 AWG soft-drawn copper wire is the most common material for ground radials. Copperweld mesh ground screens, such as are employed for lighting protection under electric power substations, are suitable for the area immediately surrounding each tower, and the radial wires can be connected to the edge of such a screen. Frequently an additional 120 radials, each 50 ft. long, are interspersed between the longer radials where a ground screen is not used. All connections within a ground system must be silver-soldered or brazed, as soft tin/lead solder deteriorates rapidly when buried.

OVERALL SYSTEM PERFORMANCE

Directional antenna pattern design and phasor design go hand-in-hand. Given a poor pattern design, no amount of clever circuit design in the phasor can provide an excellent system. Following are some considerations that can lead to an excellent antenna system.

In the Pattern Design

- Avoid closely spaced towers; 90 electrical degrees or more should be the objective.
- Avoid broadside minimums without ample spacing between towers, otherwise the driving point resistances can become very low.
- Check to see that the individual tower radiation vectors add in-phase (or very nearly in-phase) in the pattern maximum, otherwise excessive circulating currents, poor bandwidth, and pattern instability can result.
- Unless other considerations dictate, choose tower heights in the range of 100 to 130°. Shorter towers exhibit lower driving point resistances and poorer bandwidth. Taller towers may require excessively large impedance transformations in the ACU networks; this condition can add another bandwidth limitation.

In the Phasor Design

- Minimize reactive power (I^2X) in the phasor components. This will avoid high Q circuits.
- Avoid excessive use of series (trim) coils to adjust the reactance of fixed capacitors. Such coil/capacitor combinations necessarily have a steeper impedance versus frequency characteristic than a capacitor alone having the proper reactance.
- Consider alternative phasor designs. The design with the fewest parts usually has the best bandwidth unless specific broadbanding circuits are incorporated.
- Derate coil current ratings by 40% and mica capacitor current and voltage ratings by 50%. Remember that RF currents under heavy asymmetrical modulation can exceed 130% of the unmodulated current and that peak voltages can exceed 300% of the unmodulated RMS voltages.

A SIMPLE BANDWIDTH TEST

A simple test can evaluate the performance of the individual towers in an existing directional antenna system. The object is to determine the amplitude of each 10 kHz sideband component as radiated by each tower. This test is possible with any common field strength meter because the selectivity of such meters is just sufficient to resolve sideband components that are 10 kHz removed from the carrier. The procedure is as follows:

1. Keep the sampling lines terminated into the antenna monitor, but add "T" connectors to bridge off samples into a field strength meter operating as a linear tuned voltmeter. Be certain the samples do not exceed the safe voltage input for the field strength meter.
2. Modulate the transmitter 50% with 10 kHz sinewaves.
3. For each tower sample, first adjust the field strength meter gain so as to set the carrier level full scale (100%). Then tune in each sideband in turn and log each sideband amplitude as a percent of full scale. In a perfect system, this would result in sideband amplitudes equal to 25% of the carrier level in each tower.
4. Repeat the process for each tower.

In the best antenna systems, all 10 kHz sideband components will range between 23% and 27% with 50% modulation at 10 kHz. Poor systems have been observed to have individual tower sideband components as low as 5% or as high as 40% in this test.

In the main lobe of a directional antenna pattern, deficient 10 kHz sideband amplitudes do not usually result in measurable distortion because the radiation components from the individual towers add approximately in phase and the sideband deficiencies of any one tower are masked by the total radiation from all of the other towers. This test is useful in pinpointing limitations in existing systems.

Sampling Source to Feed a Modulation Monitor

The indications of a modulation monitor may be inaccurate at high modulation frequencies if the modulation sample is taken from a less than perfect antenna system. In an antenna system with a high VSWR at the 10 kHz sidebands, the voltage sampled for the modulation monitor may be higher or lower at the 10 kHz sidebands than at low modulating frequencies.

The best measure of 100% modulation is at a remote point in the main lobe of the antenna pattern. Various sampling points within the transmitter or phasor can be tried until a sample source is found that matches the percentage modulation at high modulating frequencies observed in the far field of the main lobe.

CONCLUSION

When the complexities in the design of directional antenna patterns and circuitry are considered together, the need for specialists in this field is evident. However, these chapters should have increased the reader's understanding of the complexities and thus assisted in the maintenance and improvement of AM directional antenna systems.

4.10

COMPUTER SIMULATION OF AM RADIO ANTENNA SYSTEMS

JAMES B. HATFIELD
HATFIELD AND DAWSON CONSULTING ENGINEERS, LLP, SEATTLE, WA

INTRODUCTION

Antenna systems have been analyzed and designed using computer simulation for many years by academics and those designing antennas for the military. By comparison to electromagnetic simulation of aircraft and ship structures, modeling medium wave antenna systems is an almost trivial exercise. It has been this writer's, and other consultants', experience that a relatively short time spent modeling an AM array can save a great deal of trial and error standing out in the rain and cold with two-way radios tuning up a directional array (DA).

FCC EQUATIONS AND MOMENT METHODS

The two pages of equations labeled Figure 4.10-1, published by the FCC in the *Code of Federal Regulations*, have been the basis of AM DA design, with few modifications, since the 1940's. *Equation (1)* of this Figure determines the basic pattern shape, while the other formulas deal with pattern size, vertical angle and standard pattern computations.

The term *moment methods* refers to the fact that electromagnetic fields from antennas are proportional to the area under the tower current distribution curve. The units of this area are (amperes) × (length) which is of the same form as the mechanical turning moment used in civil engineering. Moment method programs find the area under the current distribution curve by a process of numerical integration where the incremental area is the product of the tower current at a point and an incremental distance.[1]

Most moment method programs compute far field horizontal plane pattern shape by using the same mathematical expressions as FCC *Equation (1)*, although with different notation. The major difference between moment method computations and the FCC equations shown in Figure 4.10-1 is that the FCC formulas for vertical plane field and pattern size computations assume sinusoidal tower current distributions, while all moment method field and pattern size computations are derived from Maxwell's equations in integral form and are scaled to the specified input power.

The moment method equation for the horizontal plane far fields from a tower can be boiled down to:

$$E = \frac{\pi}{3}\sum i \cdot \Delta l$$

The inverse distance field, in mV/m, at 1 km, from a tower is 1.0472 times the sum of the incremental tower current moments in amp-degrees. This summation of incremental current moments over the length of the tower is usually called the *tower moment*.

Since the field from each tower in an array is 1.0472 times the tower moment of that tower, the field ratio of each tower is the ratio of the tower moment of that tower to the tower moment of the reference tower. The tower moment is a polar number, with a magnitude and angle. The phase angle of the field ratio is found by subtracting the angle of the reference tower moment from the tower moment angle of the tower in question.

PRODUCING THE CORRECT PATTERN WITH MOMENT METHOD PROGRAMS

The FCC equations use variables called *field ratios* and *current phases*. These field parameters are specified in FCC Form 301 applications and on licenses as *theoretical parameters*. The current phase, in practice, refers to the relative phase angle of the *antenna loop* current. The loop current refers to the maximum antenna current. The FCC equations are based upon the assumption that the loop current phase angle and the phase angle of the contribution of the tower to the far field are the same. It is important to note that FCC *Equation (1)* uses the field ratio as the relative contribution of the tower in question to the far field. The phase angle used in the equation is the phase angle of that far field contribution. In fact, the ratio and phase angle of the field parameters are determined by the behavior of the tower as a whole. The ratio and phase angle of the tower loop current is seldom, if ever, the same as the actual field ratio and phase angle. This difference is the reason that antenna monitor readings do not usually match the theoretical parameters. If we can compute what the antenna monitor ratios and phase angles should be we can avoid a great deal of trial and error.

The antenna system can be thought of as a black box where the inputs are voltages and the outputs are fields. The antenna system is linear, so the inputs and outputs can be related by a series of constants even if the towers are not all of the same height.

If one includes mutual impedance, the relationship between the voltage drives and fields for a two tower array is of the following form:

$$E_1 = V_{11}T_{11} + V_{12}T_{12}$$
$$E_2 = C_{21}T_{21} + V_{22}T_{22}$$

"E" represents the FCC field parameters (ratios and phases), "V_{11}", "V_{22}", are the tower base drive voltages, "V_{12}", "V_{21}", are the tower base voltages induced by the other towers, and the "T" is the constant that relate them. The "T" constants are found by shorting the respective towers to set the pertinent base voltages to zero.

Moment method computer programs devoted to broadcast antennas solve these equations for the base voltages necessary to produce the proper pattern. Base drive parameters for the correct FCC pattern can be computed by some method moment programs for top loaded or self supporting towers.

Antenna monitor ratios and phase angles can be found by calculating the ratios and relative phase angles of the computed tower currents at the location on the tower where the antenna monitor samples the current.

Some programs include a provision for current drives. If the antenna sample system is well characterized (sample line lengths and sample transformer sensitivities are known) the correct base current ratios and phases can be determined and used as base drives for computation of the pattern as adjusted.

Near-Fields and Proximity Effect

FCC *Equation (1)* calculates the pattern shape using the assumption that lines are parallel between the towers and the location where the field is being calculated. This assumption is true only at some distance from the array. Closer than this to the towers the pattern is not properly formed. The measured fields close to the antenna on null radials will not be the same as FCC *Equation (1)* would predict even though the patten is properly adjusted. This has been called the *proximity effect* and may, or may not, be an important factor depending upon array geometry.

The proximity effect occurs in an area called the *array near-field*. Accurate computations of the fields in the array near-field region can be made with moment method programs. These near-fields can then be compared to FCC *Equation (1)* fields computed at the same locations for analysis.

Field strength meters use shielded magnetic field sensing loops and convert the magnetic field component of the radiated signal to equivalent electric field units. The magnetic field is multiplied by 377 to A/m to V/m. The conversion factor of 377 only applies to plane wave fields and, in the presence of reradiating objects, may result in an improper indication of the equivalent electric field. Magnetic fields and electric fields are computed separately by the near-field computational portion of the NEC and MININEC family of programs. As a result measurement anomalies resulting from non-plane-wave conditions can be determined.

Impedance

Tower self impedances computed by method of moment program are benchmarked to values computed in R.W.P. Kings' *Theory Of Linear Antennas*. The King values for monopole antennas are based upon measurements made at high frequencies using the center conductor of a coaxial cable projecting through a copper sheet ground plane to which the outer conductor is connected. This results in resistance and reactance values that differ significantly from the resistance and reactance of a base insulated guyed tower over a radial wire ground system. The shunt capacitance of the tower base insulator can have a significant transforming effect upon high tower base impedances. Tower feed connections also add a significant inductive reactance to the tower reactance.

These defects in the computation of tower impedance can be corrected by the addition of reactance to the tower model (Figure 4.10-3) or by increasing the tower height by about 6.7%. It must be noted, however, that the fields and tower radiation efficiencies computed by moment method programs for actual tower heights are close to measured values and also to those shown by FCC graphs (see Figure 4.10-2).

Detuning Towers

When making field strength readings it is sometimes necessary to detune nearby reradiating objects which cause localized perturbations in the measured fields. The moment method programs can be used to determine whether the reradiating object affects the station's pattern or only the measurements near the reradiating object.

An effective method for detuning towers is to treat the tower as a part of the array. The field parameters for the detuned tower are set to zero as a program data input. When the computed drive voltage (and phase) are applied to the tower to be detuned, along with the proper drives to the array, and the computation is performed, the impedance of the tower to be detuned is noted. If the reactive component is approximately ten times the resistive component for the detuned tower it can be detuned by loading it with the conjugate of the computed impedance. If the computed reactance is $-j450\ \Omega$ the tower would be detuned by $+j450\ \Omega$ across its base to ground.

If the resistive part of the detuned tower impedance is large, and negative, it may, in addition, have to be loaded by a resistance in series with the detuning reactance. If the resistive component of the detuned tower impedance is large and positive it will have to be driven with a small amount of power.

The effects of the shunt capacitive reactance of the tower base insulator must be taken into account when detuning a tower. The best procedure for being sure that a tower is properly detuned is to send someone a third of the way up the tower with a current sensing device (if the tower is over about 130° tall the height will be greater, see Figure 4.10-6). The detuning reactance is adjusted until a tower current minimum is observed at the detuning point. A field set with the

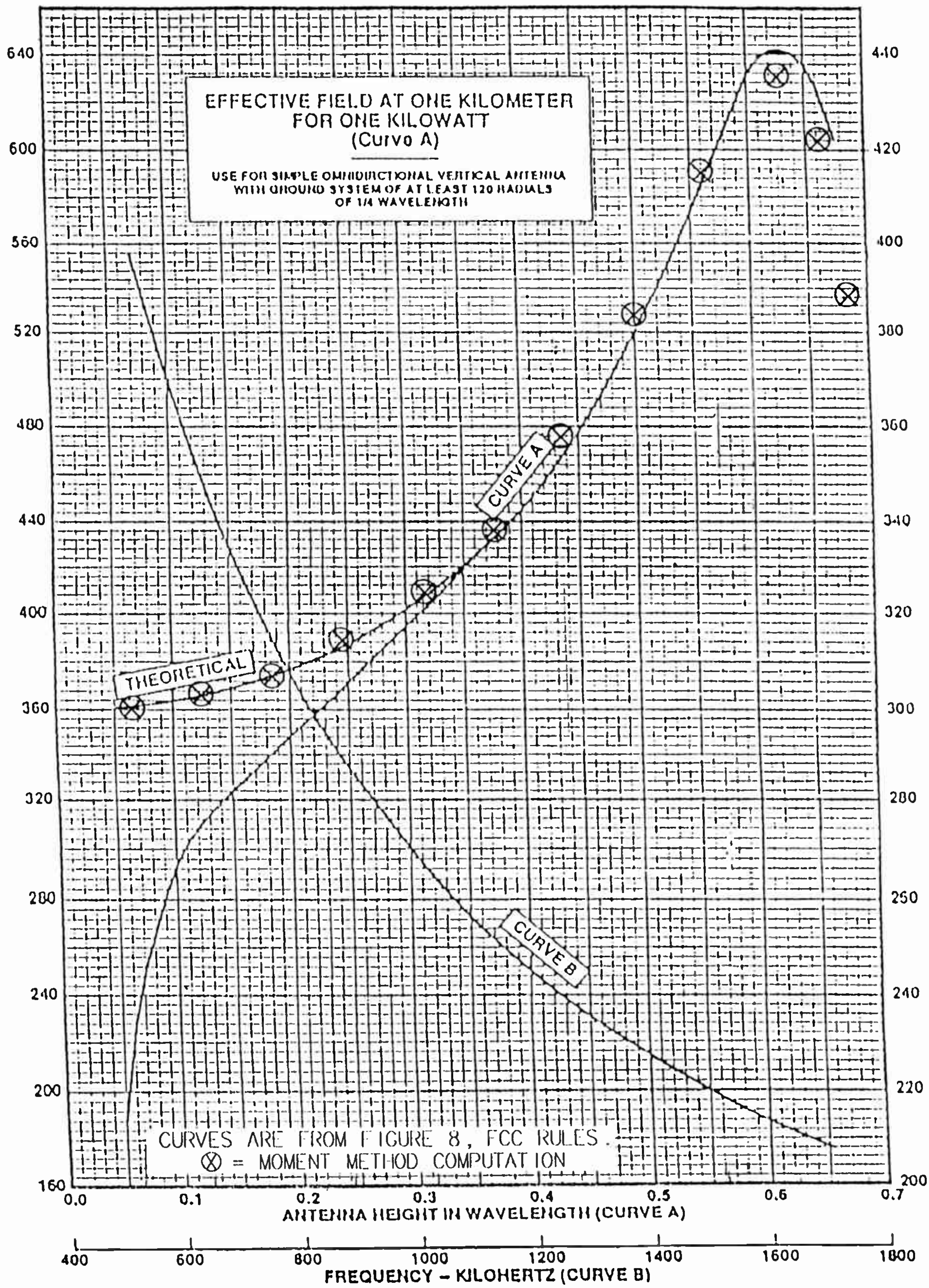


Figure 4.10-2. Lossless inverse field at one kilometer.

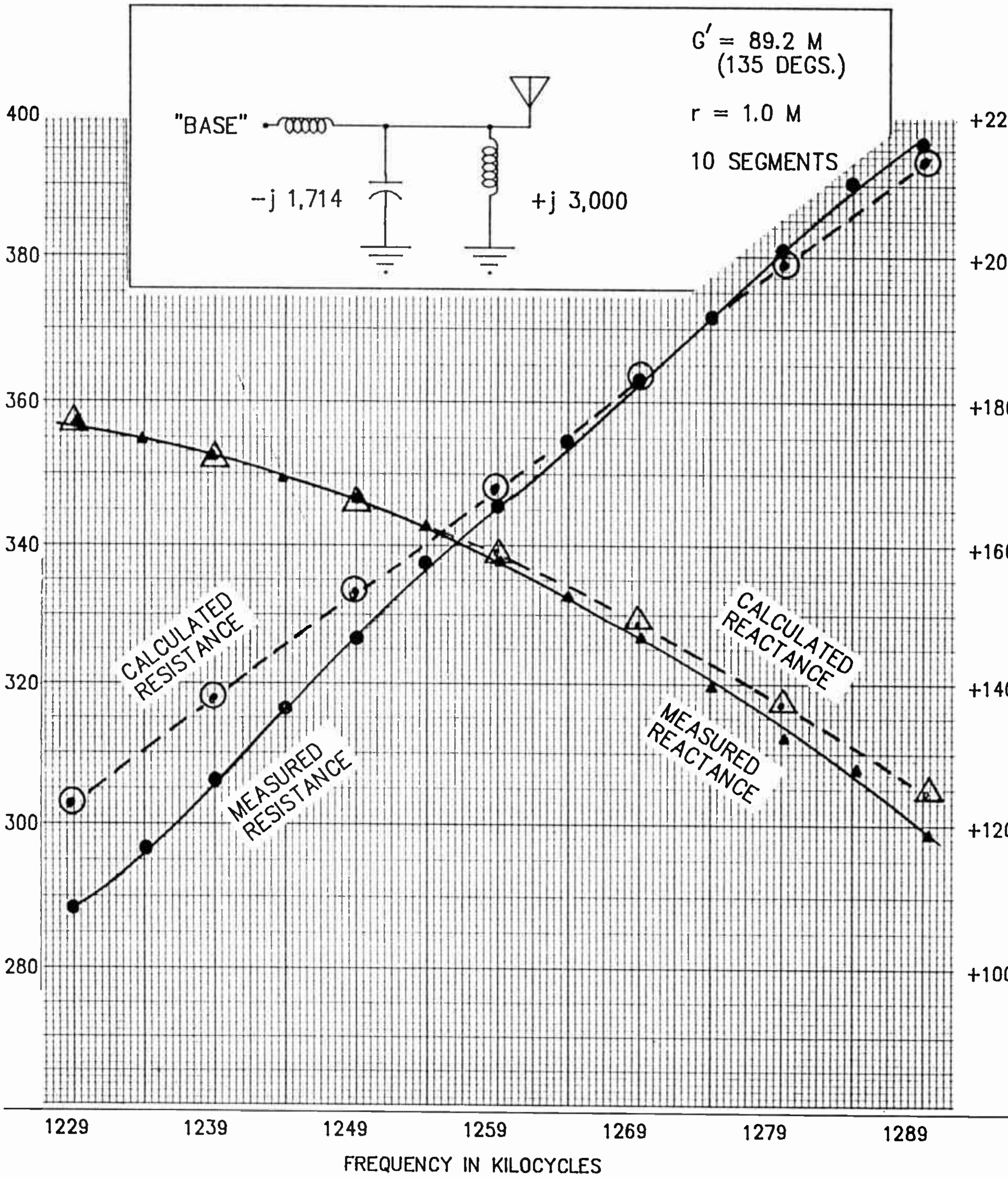


Figure 4.10-3. Rhodes, Greece tower model measurement and analysis by Ron Rackley.

loop shield shorted out through use of a screw can provide a sensitive indication of when the tower current minimum is reached.

This procedure is based upon the fact that towers are detuned in the horizontal plane by having the area under two-halves of the tower current distribution curve being equal and opposite in phase. If the current on a 90° tower is nearly sinusoidal the area under the curve is given by the cosine function. On a per unit basis the cosine is 0.5 at 30° and one-half of the area is one-third of the way up the tower.

Figure 4.10-4 shows the effects upon the nondirec-

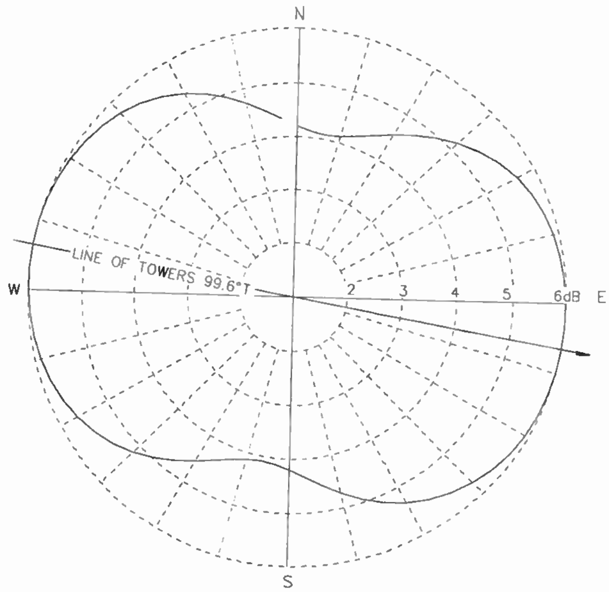


Figure 4.10-4. Non-directional patter, 3-tower array center tower driven, end tower floated, 100 pf to ground.

tional fields of the center tower of a 90° tall three tower array when the end towers are not detuned.

While detuning drastically lowers the fields in the horizontal plane it can cause large increases in vertical plane fields. Figure 4.10-5 illustrates this effect upon a four tower 110° array near a tall (5/8 wave) detuned tower. For a daytime directional, this effect may not be important, but it could have a disastrous effect upon radiation minima at night at high pertinent vertical angles.

Matching Antenna Monitor and Theoretical Field Parameters

The location on a detuned tower where the antenna current goes through a minimum and a reversal of phase angle is at the center of the area of the current distribution curve. The current at this location is proportional to the tower moment and hence the radiated field from the tower. If a sample loop is placed at this location and the sample system is the same for all towers of equal height, then the antenna monitor parameters will be the same as the theoretical field parameters.

The sample loop location for identical antenna monitor and theoretical field ratios and phase angles is shown in Figure 4.10-6 as a function of antenna height. For towers below 110° the sample loop should be placed one-third of the way up the tower. It is necessary, of course, that the rest of the sample system be identical, with equal length, same type, sample lines up to the loops and identical sample loops on all towers, with all loops oriented in the same direction.

When the towers are of different heights, the monitor ratios can be determined by taking the ratios of the computed currents one-third of the way up the towers. The monitor phase angle relative to the reference tower will be the same as the relative theoretical field phase angle.

The ratio of the current at a point on a tower to the tower moment is nearly constant for varying drive conditions. For different height towers the monitor ratio one third of the way up the tower for tower two of a two tower array is $(M_1/I_1) \times (I_2/M_2) \times F_2$. "M" refers to the tower moment for the respective numbered towers, "I" is the current on the tower one third of the way up the tower (slightly higher for towers over 130°) and "F" is the field ratio for the non-reference tower, while the monitor phase angle for the non reference tower is the same as the theoretical field phase angle. The tower moment and corresponding sample current can be calculated for each tower separately, independently of drive conditions, and the correction factor for the antenna monitor current ratio of the non-reference tower will be accurate for all field parameters.

When the antenna monitor sample loops are located at the tower height shown in Figure 4.10-6, the correction to the antenna monitor sample current ratio for

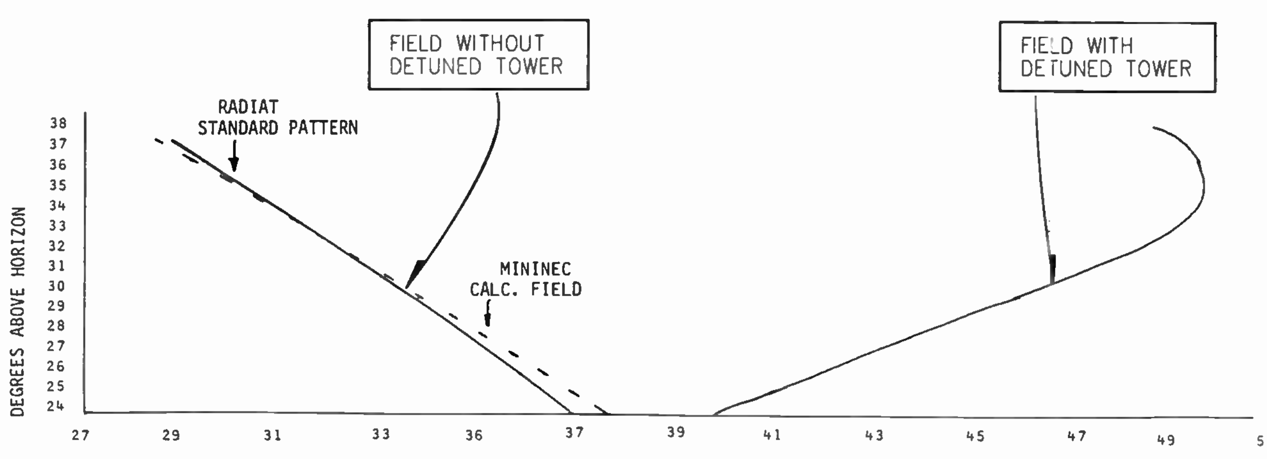


Figure 4.10-5. Effect of detuned tower on vertical angle radiation.

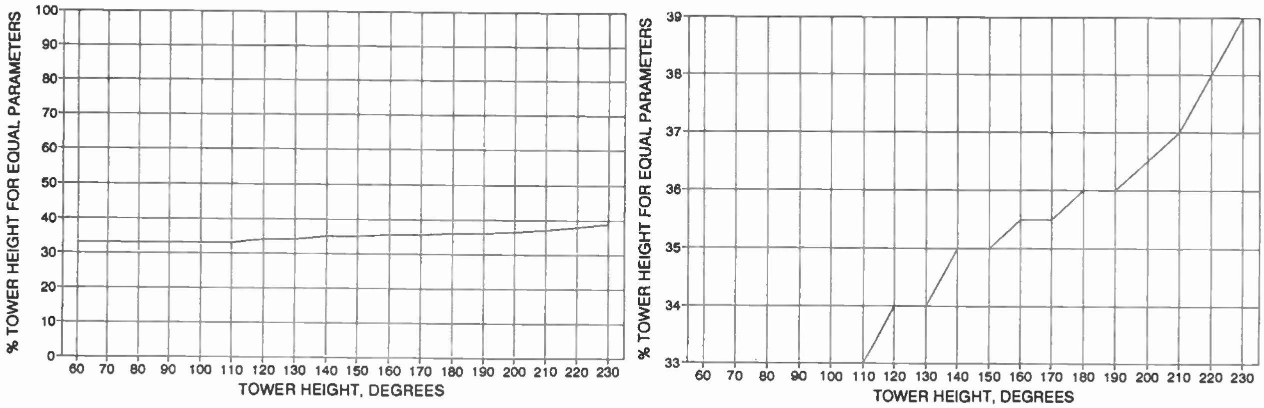
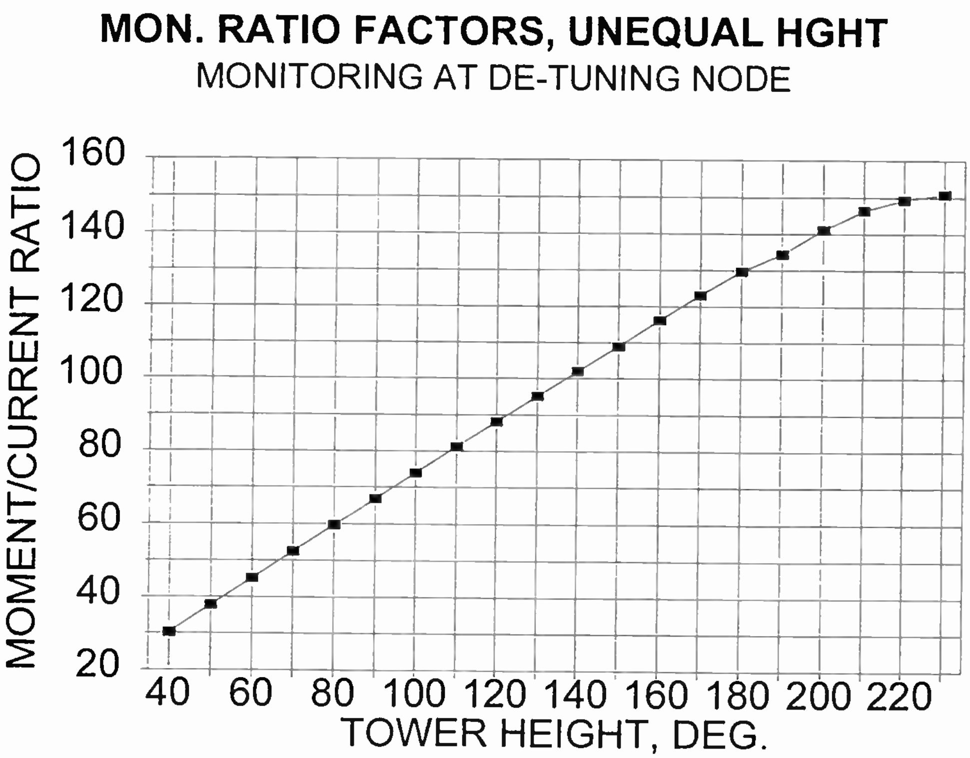


Figure 4.10-6. Percent of tower height where current ratios are same as field ratios and location of detuning current minimum.

towers whose height is different from the reference tower can be found by using the tower moment to sample current ratio factors (M/I) shown in Figure 4.10-7. To find the correct antenna monitor current ratio for a tower that differs in height from the reference tower, multiply the reciprocal of the M/I factor for tower height by the M/I factor for the tower height. This result is multiplied by the theoretical field ratio of the tower in question to determine the corrected antenna monitor current ratio for that tower.

Since the fields differ from "M" by a constant, one could, in theory, determine the correction to the monitor ratio for unequal height towers by measuring the tower currents at the sample location and the fields at the same distance from each tower (or use FCC *Figure 8* from Section 73.190, shown as Figure 4.10-2) under identical tower drive conditions, and compute the correction to the antenna monitor current ratio from the above relationship. This is not always a practical procedure, however, due to the limitations of measurement accuracy and the difficulty of isolating the tower to be measured.



HEIGHT, DEGREES	FACTOR
40	30.28
50	37.72
60	45.09
70	52.39
80	59.63
90	66.8
100	73.9
110	81.15
120	88.12
130	95.06
140	101.9
150	108.9
160	116.1
170	123.1
180	129.7
190	134.4
200	141.2
210	146.4
220	149.3
230	150.9

Figure 4.10-7. Chart and nomograph of tower height factors for correcting field ratios for unequal height towers.

Arrays Near Reradiating Objects

Figure 4.10-8 shows a typical problem that arises when an AM array is near a power transmission line. There are several effects that occur simultaneously:

- The current flowing through the transmission line towers and skywires causes large changes to the ratio of the electric and magnetic components of the radiated field of the array
- The power line may only affect the accuracy of nearby field strength readings but not the inverse distance pattern of the station
- The power line may affect both the field strength readings and the station pattern.

These effects can be separated out by modeling the power line and the AM array and then computing the near electric and magnetic fields along with the far-field pattern as influenced by the power line. A 1 Ω loss pattern can be computed by subtracting the square of the loop currents (maximum currents on the print-out) from the station input power. Corrections to the field strength meter readings can be calculated from the magnetic near-field data.

In one instance there were three tall guyed communications towers within a mile of an AM array. The array and the guyed towers were modeled and the stations pattern was scalloped as shown in Figure 4.10-9. The computed pattern did not exceed the standard pattern so no attempt was made to detune the communications towers.

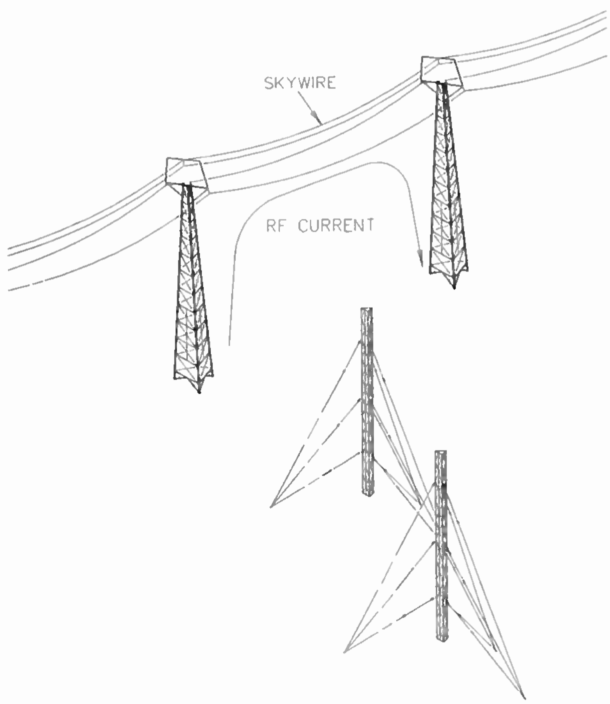


Figure 4.10-8. Two tower array with power line and skywire.

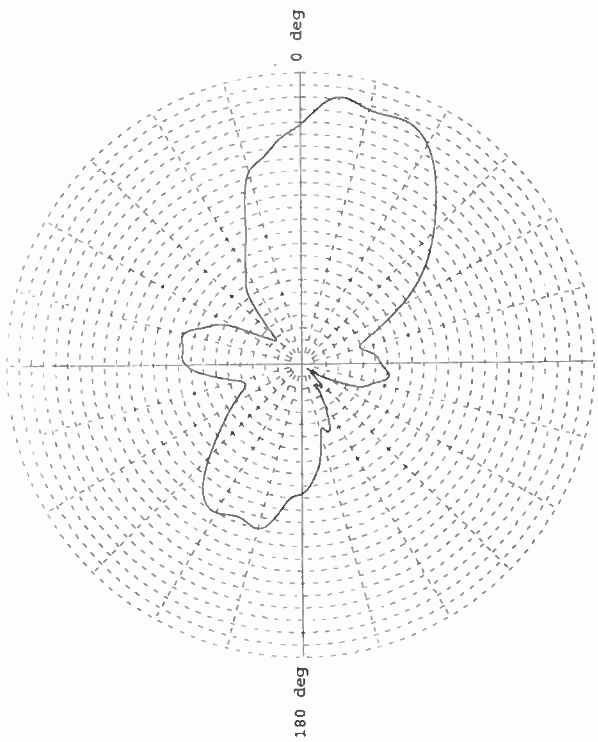


Figure 4.10-9. Example of the effect of a nearby reradiation on a DA pattern.

Top Loading

To determine the effects of modest amounts of guy wire top loading, a 76° non-directional tower was modeled as top loaded to 8° and 14°. The top loading, in electrical degrees, was simply the length of uninsulated guy wire connected to the top of the tower. Horizontal skirts were also added to the ends of the guy sections to determine their effects.

Eight degrees of top loading increased the resistive component of the base impedance by 45% while the reactance was reduced to 12% of the un-top loaded base reactance and went from capacitive to inductive. Adding horizontal skirts caused a further 16% increase in base resistance while the reactance became more inductive by a factor of five.

Fourteen degrees of top loading produces results that are, practically speaking, indistinguishable from skirted 8° of top loading with regard to resistance and reactance. Adding skirts to 14° of top loading tripled the inductive reactance and increased the resistance by 28%.

The conclusions are: top loading makes short antennas more inductive; the Q of the antenna is worse with skirted top loading and optimum with modest (8°) unskirted top loading. For all top loading from 0° to 14°, skirted or not, the inverse field at 1 km. only varied from 310 to 314 mV/m. So one can conclude that 8° to 14° of top loading has only a minor effect on tower radiation efficiency.

COMPUTING HUMAN EXPOSURE TO RF FIELDS FROM AM ARRAYS

Human exposure to magnetic and electric fields from AM towers can be computed using moment method programs. Licensed facilities in this country are restricted to a maximum power of 50 kW, so the FCC exposure limits are reached at a distance from each tower that is much less than the distance between adjacent towers. For this reason each tower can be treated as a separate case if no other sources of exposure are present at the site.

Tables 1 through 4 of distances to fences around AM towers shown in *Supplement A to OET Bulletin 65 (Edition 97-01)* were based upon moment method computations. The Commission has accepted such computations for RF guidance level determination for many years.

Figure 4.10-10 shows a computation of the electric and magnetic fields around an AM tower. The measured fields agree quite closely with the computed values. The figure also shows that one must model the tower carefully if correct results are to be achieved.

KEY TERMS

Antenna Monitor Parameters. The relative antenna current magnitudes and phase angles at a specified location on the towers.

Current Loop. The maximum current location on a tower. This is at the base of the tower for tower heights of a quarter wave or less.

Electrical Degrees. A unit of distance proportional to the free space wavelength at the frequency of interest. One wavelength is 360° and a quarter wave tower has an electrical height of 90°.

Field Parameters. The relative magnitude and phase of the contribution to the far electric field of each tower in an AM array.

Inverse Field. The electric far field of an AM antenna at 1 km that is not attenuated by earth losses and varies with distance R proportional to 1/R.

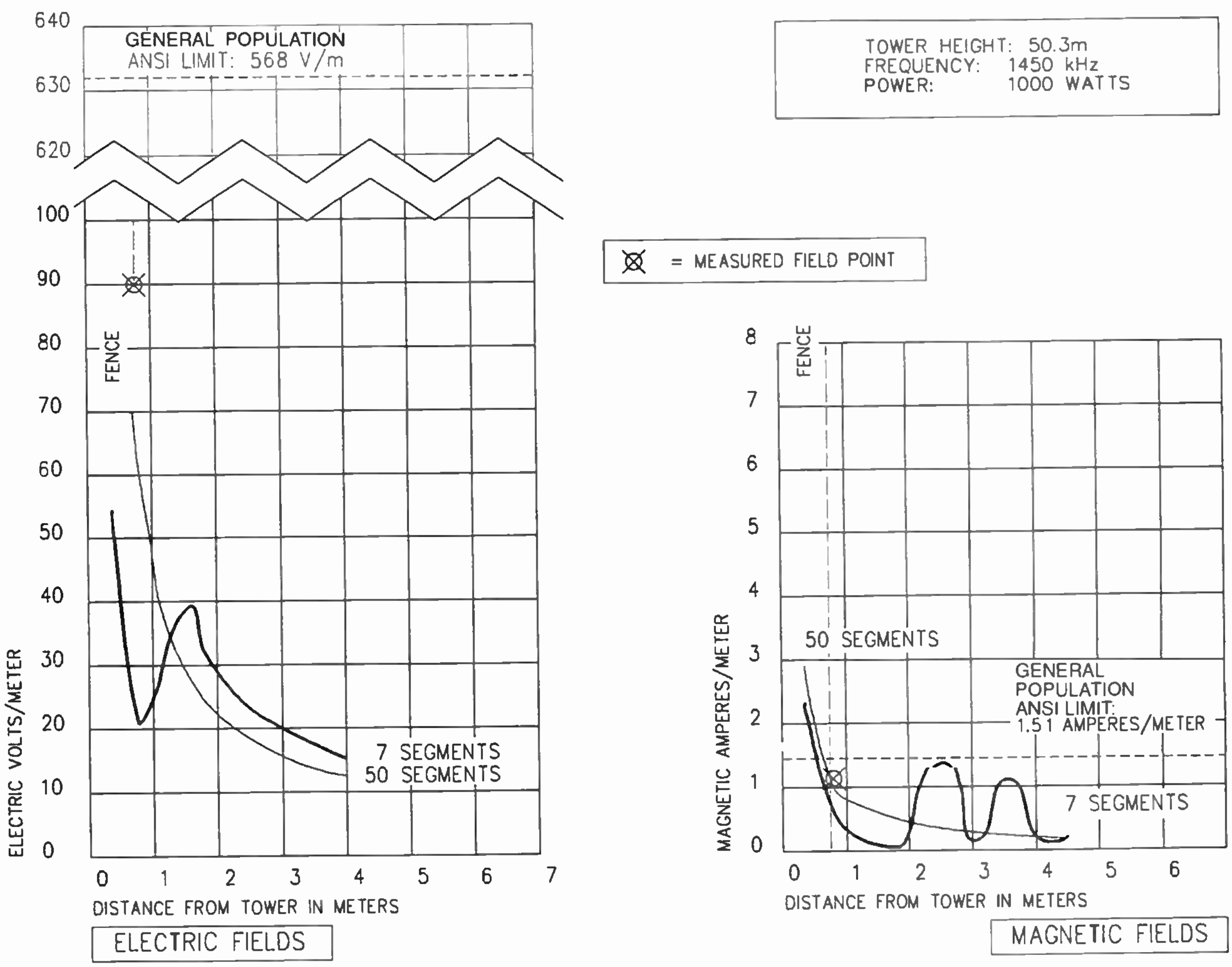


Figure 4.10-10. Electric and magnetic fields at 1 meter above ground near an AM tower.

Method of Moments. In the method of moments the integral equations relating antenna currents to radiated electric and magnetic fields is approximated by a set of linear equations. In these equations the differential distance is expanded to finite size, (segment) and the integral is replaced by a summation of a fixed number of current-segment products. When the matrix formed by these linear equations is inverted, the unknown variables can be expressed in terms of known input variables and numerically evaluated.

Sinusoidal Current Distribution. Antenna current that varies with antenna height proportional to the sine or cosine of the antenna height in electrical degrees.

BIBLIOGRAPHY

47 CRF Part 73, Subpart A (October 1, 1997).

Federal Communications Commission Office of Engineering & Technology. *Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Field,* OET Bulletin 65, Edition 97-01, Supplement A, August 1997.

The Federal Register, June 15, 41(116).

King, R. W. P., 1956. *The Theory of Linear Antennas,* Harvard University Press, Cambridge, MA.

Lahm, K., "Excitation Sampling in MW Directional Arrays," Broadcast Technology Society 44th Annual Symposium, September 1994.

Westburg, J. M., "Matrix Method for Relating Base Current Ratios to Field Ratios of AM Directional Stations," *IEEE Trans. On Broadcasting 35(2)* 1989.

Whitaker, Jerry C., "Computer Analysis of Antenna Systems," *The Electronics Handbook*, pp. 1353–1357, CRC Press & IEEE Press, 1996.

ADDITIONAL SOURCES OF INFORMATION

Short Course on Modeling Broadcast Antennas, Applied Computational Electromagnetic Society, Richard W. Adler, Executive Officer, ECE Department, Code ECAB, Naval Postgraduate School, 833 Dyer Road, RM 437, Monterey, CA 93943-5121, Phone:408-646-1111, Fax: 408-649-0300, e-mail: RWA@IBM.NET.

NEC-AM, Copyright 1988, is available from David J. Pinion, P.E. at 1202 East Pike St., Suite 1217, Seattle, WA 98122-3934, Phone: (206) 323-4631.

Phasor Professional is available from Jerry M. Westberg, 3326 Chapel Valley, Quincy IL 62301, Phone: (217) 223-6702.

MININEC Broadcast Professional for Windows is available from EM Scientific Inc., 2533 N. Carson Street, Suite 2107, Carson City, Nevada 8906-0147, Phone: (702) 888-9449, Fax: (702) 883-2384, e-mail: 76111.3171@compuserve.com, Website: http://www.sierra.net/emsci/.

RULES AND REGULATIONS

(Eq. 1)

$$E(\phi, \theta)_{th} \left| k \sum_{i=1}^{n} F_i f_i(\theta) \underline{/S_i \cos\theta \cos(\phi_i - \phi) + \psi_i} \right|$$

where:

- $E(\phi, \theta)_{th}$ Represents the theoretical inverse distance field at one kilometer for the given azimuth and elevation.
- k Represents the multiplying constant which determines the basic pattern size. It shall be chosen so that the effective field (RMS) of the theoretical pattern in the horizontal plane shall be no greater than the value computed on the assumption that nominal station power (see 73.14(c)) is delivered to the directional array, and that a lumped loss resistance of one ohm exists at the current loop of each element of the array, or at the base of each element of electrical height lower than 0.25 wavelength, and no less than the value required by § 73.189(b)(2) if this part for a station of the class and nominal power for which the pattern is designed.
- n Represents the number of elements (towers) in the directional array.
- i Represents the element in the array.
- F_i Represents the field ratio of the i[th] element in the array.
- θ Represents the vertical elevation angle measured from the horizontal plane.
- $f_i(\theta)$ Represents the vertical plane distribution factor of the i[th] antenna.

  For a typical vertical antenna with a sinusoidal current distribution:

  $$f(\theta) = \frac{\cos(G \sin\theta) - \cos G}{(1 - \cos G)\cos\theta}$$

  where G is the electrical height of the tower.

  See also Section 73.190, Figure 5.
- S_i Represents the electrical spacing of the i[th] tower from the reference point.
- ϕ_i Represents the orientation (with respect to true north) of the i[th] tower.
- ϕ Represents the azimuth (with respect to true north).

Figure 4.10-1(a). Formulas given by the FCC for determining directional antenna pattern size and shape taken from the Federal Register.

ψ_i Represents the electrical phase angle of the current in the i[th] tower.

The standard radiation pattern shall be constructed in accordance with the following mathematical expression:

(Eq. 2) $$E(\phi, \theta)_{std} = 1.05\sqrt{\{E(\phi, \theta)_{th}\}^2 + Q^2}$$

where:

$E(\phi, \theta)_{std}$ Represents the inverse distance fields at 1 km which are deemed to be produced by the directional antenna in the horizontal and vertical planes.

$E(\phi, \theta)_{th}$ Represents the theoretical inverse distance fields at 1 km as computed in accordance with Eq. 1.

Q is the greater of the following two quantities:

$$0.025g(\theta)E_{rss}$$

or:

$$10.0g(\theta)\sqrt{P_{kw}}$$

where:

$g(\theta)$ Is the vertical plane distribution factor, $f(\theta)$, for the shortest element in the array (see Eq. 2, also see Section 73.190, Figure 5). If the shortest element has an electrical height in excess of 0.5 wavelength, $g(\theta)$ shall be computed as follows:

$$g(\theta) = \frac{\sqrt{\{f(\theta)\}^2 + 0.0625}}{1.030776}$$

E_{rss} Is the root sum square of the amplitudes of the inverse fields of the elements of the array in the horizontal plane, as used in the expression for $E(\phi, \theta)_{th}$ (see Eq. 1), and is computed as follows:

$$E_{rss} = k\sqrt{\sum_{i=1}^{n} F_i^2}$$

P_{kw} Is the nominal station power, expressed in kw; see Section 73.14. If the nominal power is less than 1 kw, $P_{kw} = 1$.

$$K = \frac{(C1)(\sqrt{P_{nom}})}{rms_{hem}}$$

where:

K = the no-loss multiplying constant;

$C1$ = 244.86422 mV/m; this is the horizontal radiation from a standard hemispherical radiator in mV/m at 1 km.

P_{nom} = the nominal power in kw;

rms_{hem} = the root mean square effective field intensity over the hemisphere, which may be obtained by integrating the rms at each vertical elevation angle over the hemisphere. The Commission's computer performs the integration using the trapezoidal method of approximation:

$$rms_{hem} \cong \sqrt{\frac{\pi\Delta}{180}\left[\frac{rms_0^2}{2} + \sum_{m=1}^{l} rms_{m\Delta}^2 \cos m\Delta\right]}$$

where:

Δ = the interval, in degrees, between the equally spaced sampling points at the different vertical elevation angles θ;

m = integers from 1 to I, which give the elevation angle θ in degrees when multiplied by Δ;

l = one less than the number of intervals; it is equal to $90/\Delta - 1$;

rms_θ = the root mean square field intensity at the specified elevation angle θ:

$$rms_\theta = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} F_i f_i(\theta)\, F_j f_i(\theta) \cos\psi_{ij} J_s(S_{ij}\cos\theta)}$$

where:

i = i_{th} tower;

j = j_{th} tower;

n = number of towers in the array;

F_i = field ratio of the i_{th} tower;

$f_i(\theta)$ = vertical radiation characteristics of the j_{th} tower;

F_j = field ratio of the j_{th} tower;

$f_j(\theta)$ = vertical radiation characteristics of the j_{th} tower;

ψ_{ij} = difference in the electrical phase angles of the currents in the i_{th} and j_{th} towers in the array;

S_{ij} = spacing in degrees between the ith and jth towers in the array;

$J_o(S_{ij}\cos\theta)$ = Bessel function of the first kind and zero order of the apparent spacing between the i_{th} and j_{th} towers.

Next, the no-loss loop current (the current at the current maxima) for a typical tower is computed:

$$I_i = \frac{KF_i}{(C2)(1 - \cos G_i)}$$

where:

I_i = the loop current in amperes in the i_{th} tower;

K = the no-loss multiplying constant computed above;

F_i = the field ratio for the i_{th} tower;

$C2$ = 37.256479; this was derived in Constants for Directional Antenna Computer Programs;

G_i = the height, in electrical degrees, of the i_{th} tower.

Figure 4.10-1(b). Formulas given by the FCC for determining directional antenna pattern size and shape. Taken from the Federal Register.

Note: If non-typical towers are used; different loop current equations may be required. If the tower is less than 90 electrical degrees in height, the base current is computed by multiplying the sine of the tower height by the loop current. Using the no-loss currents, the total power loss would be:

$$P_{loss} = \frac{R}{1000} \sum_{i=1}^{n} I_i^2$$

where:

P_{loss} = the total power loss in kw;
R = the assumed resistance in Ω; for standard pattern calculations this would be at least 1 Ω;
i = the i_{th} tower;
n = the number of towers in the array;
I_i = the loop current (or base current if the tower is less than 90 electrical degrees in height) for the i_{th} tower.

Finally, the multiplying constant must be adjusted to change the assumption from nominal power being radiated to nominal power being the input power to the array prior to taking account of the assumed loss resistance:

$$K_\Omega = K\sqrt{\frac{P_{nom}}{P_{nom} + P_{loss}}}$$

where:

K_Ω = the multiplying constant after adjustment for the assumed loss resistance;
K = the no-loss multiplying constant;
P_{nom} = the nominal power in kw;
P_{loss} = the total power loss in kw.

The multiplying constant $K\Omega$ is the used to compute the theoretical pattern used in generating the standard pattern.

[FR Doc. 76-17167 Filed 6-14-76 8:45 AM

Figure 4.10-1(c). Formulas given by the FCC for determining directional antenna pattern size and shape taken from the Federal Register.

4.11
MAINTENANCE OF AM BROADCAST ANTENNA SYSTEMS

EDWARD EDISON, P.E. & GERHARD J. STRAUB, P.E.
HAMMETT & EDISON, INC., SAN FRANCISCO, CA

INTRODUCTION

Many stable directional arrays, being electrically passive devices, operate for 10 or more years without any readjustment. However, maintenance attention will be needed to overcome the normal effects of age and deterioration (routine maintenance), to restore proper operation after catastrophic failures (such as caused by lightning), and to cope with new sources of reradiation that may develop within the environment in which the antenna system operates. Infrequently, readjustment and relicensing with new pattern parameters may be required.

ROUTINE MAINTENANCE

Routine mechanical and electrical maintenance is required to offset the effects of age and deterioration. Mechanical maintenance includes the attention necessary to keep the moving parts functioning and the tower painted and in good repair. Electrical maintenance may involve replacement or recalibration of base current meters; checks on the continuing proper performance of the phasor components, transmission lines, and ground system; and infrequent readjustment of the antenna phase and ratio parameters as necessary to ensure operation within licensed limits with the proper directional antenna pattern.

Mechanical Maintenance

The moving components within an antenna system may include RF relays, meter switches, dial drives and the rollers on variable coils.

The contacts on RF relays that carry substantial currents may eventually wear out and require replacement. It is vital that such relays be interlocked so they cannot be switched while transmitter power is applied. Relay maintenance problems are reduced when vacuum relays are employed.

Meter switches, that disconnect thermocouple-type ammeters from the circuit when not being read, are subject to considerable mechanical wear and tear. Occasional lubrication and tightening of the components are necessary if the parts are to function properly. Replacement of thermocouple ammeters with modern transformer type meters eliminates the need for RF meter switches and their associated maintenance requirements.

Dial drives on variable coils and capacitors, once properly adjusted, require little attention except infrequent lubrication. The thrust bearings on vacuum variable capacitors are factory lubricated and should not require maintenance attention. However, the rollers on variable coils are often a source of mechanical difficulties (particularly in older coils) and periodic maintenance is required if they are to operate smoothly and without excessive arcing. Networks that are designed to utilize variable vacuum capacitors instead of variable coils are superior in this regard.

Electrical Maintenance

Thermocouple RF ammeters that measure base currents and common point currents are a frequent problem due to failure or changes in calibration. If such meters cannot be replaced with modern transformer type units, an intermediate improvement can be effected by substituting make-before-break meter jacks at the base of each tower (in lieu of the original permanent thermocouple meters and associated protective switches). The meters can then be stored in a dry environment and taken to the towers only when base current readings are required. The effects of calibration errors are reduced if the same meter is used to measure all base currents (provided all currents fall within an acceptable range on the meter scale).

Thermocouple ammeters can be calibrated with 60 Hz current. A useful procedure is to remove all such meters from the system and to connect them in series for testing. A high quality dynamometer or soft iron type meter should also be included as a calibration standard. A filament transformer having adequate secondary current rating can be used to drive current through all of the meters connected in series. Current amplitude is easily controlled by supplying the filament transformer from an adjustable ac supply such as a Variac.

Modern current transformer RF ammeters cannot be calibrated with 60 Hz and, if defective, must be returned to the manufacturer for repair and recalibration. The transformer, meter and interconnecting cable are calibrated as a unit. The cable length should not be changed without recalibration by the manufacturer.

Occasionally loose connections or deteriorating phasor components will produce abnormal heating. Failures from these causes can be anticipated by simply feeling all RF connections and components for warmth immediately after sign-off.

Transmission Lines

Air or foam dielectric lines in continuous lengths rarely develop troubles, but pressure should be maintained on air dielectric lines in order to prevent the accumulation of moisture. Breaks in the outer conductor of foam filled lines that are buried may cause obscure symptoms and prove difficult to locate. In extreme situations, a time-domain reflectometer is the best tool to pinpoint the problem.

Ground System

Significant deterioration of a ground system can manifest itself in two ways. The first indications may be unusual changes in the phase and ratio parameters between wet and dry conditions. These changes depend upon the nature of the directional array and the character of the ground. When ground systems are installed in shallow tidewater or saltwater marshes, the inherent ground conductivity is so good that even drastic changes of the ground system may produce negligible effects.

A second, more obvious symptom of a deteriorating ground system, is a reduction in antenna radiation and a reduction in base currents. It is not uncommon in older arrays to find that the absolute value of each base current has fallen off since the time of the initial installation, even though the base current ratios remain correct. With the available common point power remaining constant, any increase in ground system losses must necessarily reduce the base currents.

It is essential that all joints in a ground system be made by brazing or silver soldering. Soft solder consisting only of lead and tin deteriorates quickly when buried.

The existence and continuity of buried ground radials is best checked by using an underground cable locator of the type used by utility companies to locate buried cables and pipes. These devices consist of a transmitter, that can be located at the tower base to place a modulated low-frequency signal on all connected ground wires and a receiver that detects the unique modulation. While in use, the receiver can be carried in a circle with a radius of 100 ft or more about each tower, so that the radial wires that are intact can easily be detected and counted.

Alternatively, if a cable locator is not available, a shielded pickup loop held near the ground and connected to the external input of a field strength meter may suffice to detect the current in individual radial ground wires. Even when a radial is broken, the current beyond the break tends to concentrate in the wire instead of the adjacent earth. A most definitive check is to uncover the distant ends of adjacent radials and check for continuity between them with a dc ohmmeter.

Towers

Towers require occasional attention to ensure their structural integrity. Periodic inspections by an experienced tower contractor are desirable. Guyed towers should be checked to confirm that they are straight and plumb and that the guy tension is correct. Towers with tubular legs can rust from the inside out; a close inspection is necessary to detect this condition while repairs are still feasible. Towers with solid legs avoid the hazards of internal rust. Galvanized towers must be painted only if required for aeronautical safety. Nongalvanized towers should be avoided, not only because rusting makes the maintenance problem more severe, but also because of their inferior electrical conductivity.

Lightning strikes and target shooters may destroy guy wire insulators. Such failures can usually be discovered with the aid of field glasses or a telescope. The electrical field gradient can be very high adjacent to a tower, particularly near the top. Arc overs and corona can be avoided by cascading guy wire insulators (johnny balls) immediately adjacent to the tower. Under severe conditions, fiberglass rod insulators or dielectric guy cables are a preferred solution. In an existing system with marginal insulation, any dirt or salt deposits on the insulators may cause corona and arc overs. Washing the insulators and coating them with silicone grease may provide a temporary remedy.

SUDDEN COMPONENT FAILURES

Lightning strikes are the most common cause of component failure in a well designed antenna system. Lightning that strikes a tower should be conducted directly to ground through the ball gap that is typically part of each tower installation. However, lightning has been known to destroy components in what appear to be well protected systems.

Damaged components can most usually be located by a careful visual inspection. Lightning currents flowing through adjacent coil turns tend to collapse them; therefore, coils should be inspected to confirm that all turns are in their original uniform alignment. The forces generated by lightning traveling through the coil strap and clip on a ribbon coil may dislodge the clip and leave the strap hanging in nearly its original position. Careful visual inspection is needed to locate dislodged coil clips.

Although vacuum capacitors are usually self healing if subjected to minor over voltages that cause internal arcing, the extremely high currents from a lightning strike can melt the plates together and short the capacitor, necessitating replacement.

If lightning damage is a recurring problem that warrants additional protection, the following steps are suggested:

1. Check the ball gap at the tower base. The balls should be arranged side-by-side so that dirt and

water will not bridge the gap with the first rain of the season. The balls should be on the same radial line through the tower axis because towers tend to twist in high winds; this twisting could otherwise change the gap between the balls. The gap should be as narrow as possible.

2. Provide a two-or-three turn choke coil in the RF feed from the antenna coupling unit (ACU) to the tower if one is not already installed. The coil can be from 5 to 10 in. in diameter depending on the station power level and the feedline size. Its function is to present a high impedance to lightning strikes so they will jump the ball gap rather than enter the ACU. Readjustment of the output arm of the matching network will be required to offset the added inductance of the lightning choke.
3. Provide a horn type lightning gap within the ACU similar to the design shown in Figure 4.11-1. The relatively clean, dry and insect-free environment inside the ACU permits a tighter gap setting and increased lightning protection than can be accomplished with an exterior gap. The path for lightning currents from the tower to the gap within the ACU should be very direct with minimum inductance and no sharp bends. Similar lightning gaps also can be installed at both ends of all transmission lines to help prevent lightning currents from getting into the phasor.
4. Adjust each gap individually by reducing the spacing until it arcs on the heaviest modulation peaks. The spacing should then be doubled for the permanent adjustment. This will be a much smaller gap spacing than is usually encountered on external gaps at the tower base.
5. Check to see that there is a dc path from each tower to ground, either through a phase sampling isolation coil, a separate static drain coil or resistor, a spare winding on a tower lightning choke, or the network circuitry within the ACU itself. Without a static drain, static charges may build up on a tower and then discharge with a small arc that may be sufficient to trip protective circuits and take the transmitter off the air.

ENVIRONMENTAL CHANGES

The ideal environment for a directional antenna system is a flat plane of high conductivity with no tall structures in the general vicinity, such as buildings, power line towers, water tanks or oil wells. Such structures (or any metallic objects of substantial vertical dimension) will have currents induced in them. These currents are dependent upon the size and shape of the structure, the frequency, and the ambient field strength with which they are illuminated. The currents in such objects make them parasitic antennas. The resulting

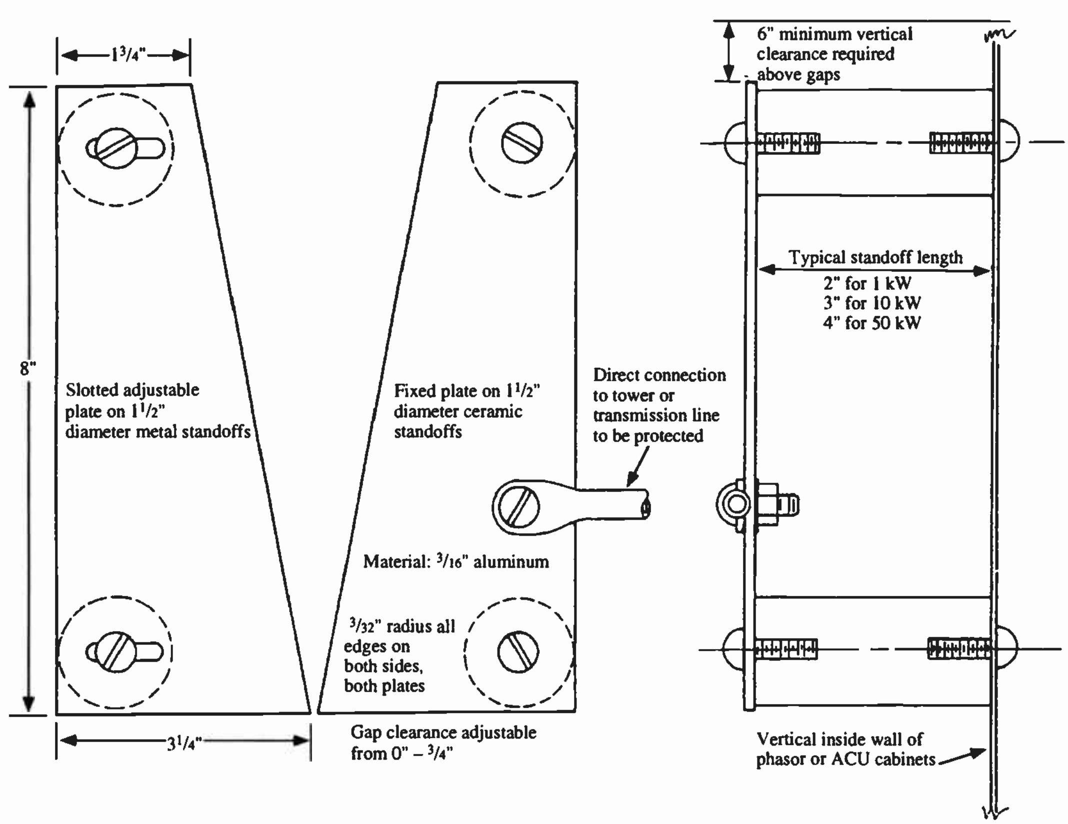


Figure 4.11-1. Suggested lightning gap design.

reradiation can severely distort the pattern of a directional antenna array.

Reradiating objects located in the main lobe of an antenna pattern are of greater consequence than those located in or near pattern minimums where their illumination is substantially less. The reradiation from reasonably slim and electrically short objects (where the overall height is substantially less than a quarter wavelength) can be calculated with useful accuracy by means of the approximate formula,

$$|E_r| = \frac{|E_i| \left[\frac{\cos(1.2\,hf) - 1}{\cos(1.2\,hf)}\right] h}{2000\left[\left(ln\,\frac{2h}{a}\right) - 1\right]}$$

Where: E_r = Reradiated field in mV/m at 1 km
E_i = Incident field in mV/m at location of reradiator
h = height of reradiator in meters
a = effective radius of reradiator in meters
f = frequency in MHz
ln = natural logarithm
cos = cosine of angle in degrees

Reradiation can also be calculated using commercially available software.

If feasible, a more accurate assessment of the reradiation can be obtained by measuring the RF current flowing in the reradiating object. If the object is reasonably thin (such as a guyed communications tower, a cable television headend tower, or the four individual legs of self-supporting power line towers) measurement of the current flowing from the tower to ground is not difficult. For this measurement, a toroidal pickup coil formed by the spiral steel wire in a length of clothes dryer exhaust hose (or even better, some types of vacuum cleaner hose) is useful. The hose section should be long enough to encircle the tower or current carrying conductor to be measured and to have its ends attached to a field strength meter operating as a linear voltmeter. Such a pickup coil can be calibrated by encircling a conductor carrying a known current, such as the feed to the base of a radio tower.

It is not necessary to measure the current distribution throughout the height of electrically short objects. The current distribution can be assumed to be triangular, decreasing to zero at the top of the tower if it is freestanding; or the current can be assumed to be uniform throughout the entire height of the tower as a worst case assumption where power line towers exhibit heavy toploading due to the capacitance to ground of the conductors they support. Once the current is measured the reradiation can be calculated by relating the *ampere feet* of current in the reradiator to that in a quarter-wave tower at the same frequency. For example, a quarter wave tower at 1,000 kHz is 246 ft high, has a base resistance of approximately 50 Ω, and a base current of about 4.5 A for 1 kW. With sinusoidal current distribution on the tower, the average current throughout its height would be 64% of the base current or about 2.9 A. Thus, at 1,000 kHz about 700 A-ft of current would produce about 190 mV per meter of radiation at one mile or 300 mV per meter at one kilometer. This corresponds to a radiation (or reradiation) in the order of 0.5 mV per meter at 1 km for each A-ft of current in the reradiator. This rule-of-thumb as to the resulting reradiation from short towers is proportional to frequency when applied to channels removed from 1,000 kHz.

CELLULAR AND PCS TOWERS

Perhaps becoming the most prevalent potential reradiating structures these days are the ubiquitous cellular and PCS towers. The FCC has codified the requirement for these facilities to notify and to conduct measurements on AM radio stations located near potential tower sites to ensure that the antenna pattern of the AM radio station is not affected by parasitic radiation from the cellular or PCS tower. The FCC requires that directional AM stations within 3 km and nondirectional AM stations within 1 km be notified. Additionally, measurements must be conducted on the AM station both before and after construction of the new tower to ensure that no adverse effects have occurred. If there have been any significant effects, the burden is on the cellular or PCS provider to remedy the problem, including detuning if necessary.

Measurements on the AM station may take a number of forms. FCC policy requires a partial proof on each pattern of a directional station, although fewer measurements may be acceptable if the station and the cellular provider agree. It is important to note that this requirement implies required cooperation on the part of the AM station, including operating with nighttime pattern during the day in order to allow field strength measurements to be made with the absence of skywave and in accordance with FCC policy for AM field strength measurements. The FCC has intervened on occasion and reminded the reluctant AM station that it has the authority to request a partial proof-of-performance from the station at any time.

Measurements on nondirectional stations should consist of approximately ten points between 3 and 16 km from the station. The measurements should be made on eight evenly spaced radials beginning with true north. The first set of measurements, for both directional and nondirectional stations, are made prior to the new tower being erected. The second set of measurements should be made after the tower has been erected and all antennas and appurtenances have been installed. The measured field strength before and after construction are then compared to look for significant changes. To avoid potential environmental changes, it is desirable to conduct the before and after measurements with as little time between them as possible. A new tower constructed very close to a nondirectional station may also affect the base impedance of the tower. In such cases, the base impedance may need to be remeasured to ensure accurate power measurement.

Even with the proliferation of these towers, there are relatively few cases where the new tower is really a problem. Most of these towers are electrically short and are such inefficient radiators that they cause no problems. For example, a typical 60 ft monopole is approximately 37 electrical degrees tall at 1700 kHz and proportionately shorter at lower frequencies in the band. It is nearly impossible for such a tower to have a significant effect on a nondirectional station at any reasonable distance from the AM tower. Such a tower located very close to a directional AM array with a very deep pattern minimum, if the tower is located in the main lobe of the AM pattern, may be able to cause some measurable effect. The same tower located in the pattern minimum would have little effect, since it is not heavily illuminated by the AM radiation.

By all means, AM broadcasters should keep diligent watch for new towers near their transmitter sites and have these evaluated by a competent engineer. However, unless the tower is relatively tall, in terms of wavelength at the AM frequency, unless the tower is located close to the AM array, or unless the AM pattern has very deep minima, it is unlikely that the tower will cause a significant effect. In cases where there is a significant effect, simple detuning measures described previously will usually remedy the problem.

RADIATION HAZARDS

In 1986, the FCC promulgated rules limiting human exposure to radio frequency fields at all AM, FM, and TV broadcast sites. In 1997, these rules were further revised. This subject is covered in detail in Chapter 9.4, *Non-Ionizing Radiation.* At AM broadcast frequencies the potential for human exposure in excess of the radiation limits exists only in the immediate vicinity of radio tower bases and in the vicinity of phasors and antenna coupling units that are not in fully enclosed cabinets. Phasors and ACUs employing open panel construction can exceed exposure limits at distances up to about 5 ft depending upon the station power level. Metallic fences surrounding each tower base form excellent shields at AM broadcast frequencies and greatly reduce the area exceeding exposure limits when compared to unprotected towers or towers within wooden fences. One useful arrangement when weatherproof ACUs are employed is to mount them so that their locked front door is in the plane of the fence surrounding the tower and is accessible without entering the fenced enclosure.

Compliance with radiation hazard limits effectively prevents maintenance or painting of a tower while it is in use. Therefore it is now highly desirable to provide for nondirectional operation on either of two towers so that every tower can be deenergized in turn when required for maintenance without taking the station completely off the air.

Although the FCC has not adopted the induced or contact RF current limits of any standard, there is the potential for creating a situation that would exceed recognized current limits near AM facilities. Tall, ungrounded, conductive structures near an AM transmitter site could have substantial current induced by the AM radiation. A person contacting the structure and ground could draw significant current from the structure. With high powers, it may even be possible to create an arc and receive an RF burn.

Large cranes are the most troublesome, as the crane, boom, and cable create a large loop antenna capable of intercepting significant energy from a nearby AM station. Contact currents on large cranes near a 50 kW facility were measured to exceed one ampere in at least one case and significant arcs could be drawn from ungrounded and uninsulated hooks. It is important to realize that this is not only a high power concern, as currents in excess of the applicable ANSI standard were measured on a large crane over 800 ft away from a 1 kW facility. While siting near cranes is not usually desirable because of the pattern distortion that they may create, if it is necessary to locate a facility near large cranes, or other objects, it may be necessary to utilize insulated hooks or other mitigation techniques on the crane or other structure.

READJUSTING AN ARRAY

Any readjustment of a directional antenna array should be undertaken only by personnel who have a basic understanding of directional antenna theory and practice. At the minimum, the material in the preceding two chapters should be read and understood. Before any array adjustments are attempted, considerable information needs to be collected and analyzed.

Four conditions for proper array performance are explained below. All of these conditions should be met simultaneously in normal operation.

- The antenna monitor indications should agree with the station license within the FCC tolerances of 5% for tower current ratios and 3° for phase. A special case is a *critical* array which has been assigned tighter tolerances by the FCC.
- Changes in base current ratios should be consistent with changes in antenna monitor ratios. Although the base current ratios and antenna monitor ratios are rarely in exact agreement (and will differ greatly in arrays with unequal height towers), these ratios should change by essentially the same percentage for small changes in the actual base current ratios. In other words, when the monitor indications match those shown in the license or the most recent proof, the base current ratios should also match what is shown in the same license or proof. Discrepancies are an indication of changes in the sampling system or in the base current metering and should be investigated.
- The reradiation environment must be essentially unchanged from that in which the array was last in proper adjustment. Any substantial structures such as power line towers or steel frame buildings recently constructed in the vicinity of the array, particularly in the main lobe of the pattern, can cause trouble.

Grounded structures approaching a quarter-wave high can distort critical antenna patterns even if located one or two miles away in the main lobe of radiation.

- The monitoring points should be within the limits shown on the station license. A high monitoring point may result from a changed environment due to the construction of buildings or power lines near the monitoring point. If a high monitoring point does not correlate with observed changes in antenna monitor parameters, the point is suspect and should be checked by measuring and analyzing 10 or more of the other points on the same radial that had been measured and recorded in the most recent proof or partial proof. Readjustment to reduce excessive radiation is indicated only if the analysis of 10 or more points confirms that the radiation in that direction is in fact in excess of the standard pattern limit.

SOIL CONDUCTIVITY

The analysis of a radial may be misleading due to changes in soil conductivity. Increased soil conductivity during wet seasons may cause substantial increases in signal strength at distant measuring points when compared to the same measurements during dry seasons. This phenomenon varies with distance, with season, and with differing geographical regions. These effects may not exist in some regions, but in extreme cases may show 2:1 changes in field strength at a distance of 20 miles. When analyzing field measurements, soil conductivity variations can be eliminated by comparing the present ratio of directional to nondirectional field strength at each measurement point to the same ratio in previous complete or partial proofs. Because new nondirectional measurements may be needed at a future date, it is important to incorporate in the phasing system a switching capability to permit reversion to nondirectional operation at those stations which do not normally have a licensed nondirectional mode.

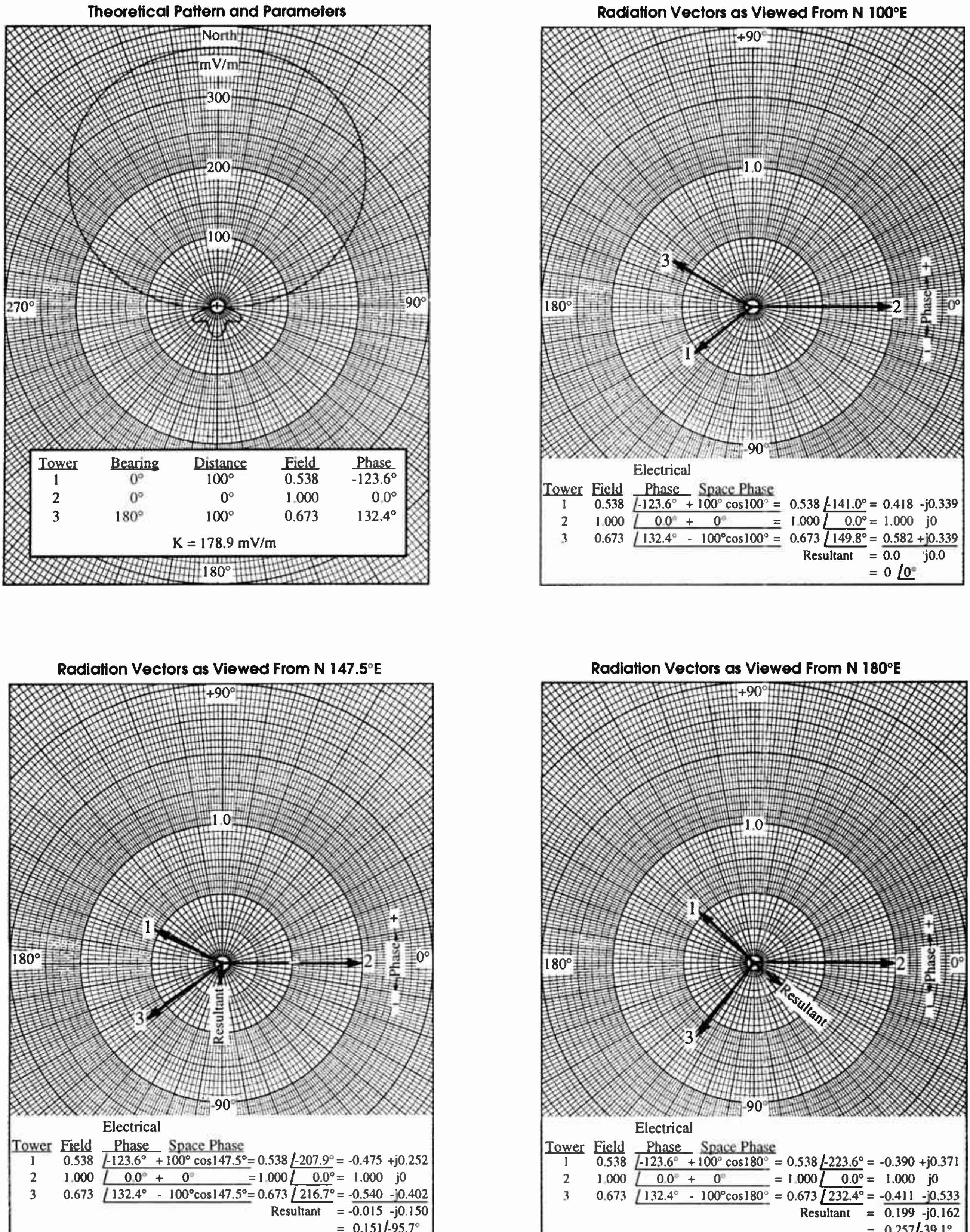
READJUSTMENT DATA AND ANALYSIS

Directional antenna arrays vary so much in pattern shape, suppression requirements, and phasor circuitry that precise readjustment instructions cannot be described. If possible, the consulting engineer who adjusted the array before its last proof of performance should be consulted on the problem. The following guidelines should prove helpful:

- Carefully log all dial settings and all coil tap locations before making any changes. Keep in mind what parameter you are trying to change before twisting any knobs or changing any coil taps.
- Make changes only in small increments, typically never more than about 1° or 1% in phase or ratio.
- Keep a step-by-step record of each change as it is made so that the array can be restored to intermediate or initial conditions, if necessary.
- Remember that even small adjustments of phase or ratio can have a measurable effect on the common point resistance. The power fed to the antenna system is only correct if the common point resistance and current are essentially correct. A permanently installed common point operating impedance bridge is a convenience not only for keeping track of the common point impedance during adjustments but also for continuing confirmation of the licensed common point value in day-to-day operation.
- Construct vector diagrams for each important direction. Samples of such diagrams are shown in Figure 4.11-2. The phase relationships must include both the electrical phase of the currents fed to the towers and the space phase component for each tower relative to the reference tower as viewed from each direction. Such diagrams will assist in visualizing what effect any contemplated changes in phase or ratio for any tower will have on the resultant field in each direction. The vector diagram amplitudes may be in terms of the individual tower field ratios (such as in Figure 4.11-2) or in terms of the theoretical unattenuated radiation. The theoretical unattenuated radiation from each tower (expressed in millivolts-per-meter at one km or one mi) is obtained by multiplying the theoretical pattern field ratio for each tower by the pattern constant K, which is also usually shown in each theoretical pattern.
- If the pattern includes essentially zero minimums (which need to be as deep as possible), make *talkdown* adjustments at a series of the most distant points on the null radial from which adequate two-way communications can be obtained. At each talkdown location, adjust the phasor for any combination of tower currents and phases that gives the deepest possible minimum. (A deep minimum can usually be confirmed by rotating the field strength meter 90° on a vertical axis and noting that the signal received from scattered reradiators is stronger than that received from the station when the meter is in its normal measurement orientation.)

Construct the vector resultant for each talkdown on a vector diagram similar to that shown in Figure 4.11-2, but use the antenna monitor indications of phase and ratio instead of the theoretical parameters. The vector resultant will probably not be zero because of errors in the antenna monitoring system, non sinusoidal current distribution on the towers, and reradiation from objects external to the array. However, the resultants so plotted from a series of talkdowns along a radial will enclose an area in which the optimum adjustment has its vector resultant at its center. Knowing the desired resultant for each critical radial direction, one may then be able to infer a set of phase and ratio parameters to satisfy conditions in all directions simultaneously.

For simple arrays, a pragmatic approach that often yields effective results, is to station observers with



Theoretical Pattern and Parameters

Tower	Bearing	Distance	Field	Phase
1	0°	100°	0.538	-123.6°
2	0°	0°	1.000	0.0°
3	180°	100°	0.673	132.4°

K = 178.9 mV/m

Radiation Vectors as Viewed From N 100°E

Tower	Field	Electrical Phase + Space Phase		
1	0.538	/-123.6° + 100° cos100° =	0.538 /141.0° =	0.418 -j0.339
2	1.000	/ 0.0° + 0° =	1.000 / 0.0° =	1.000 j0
3	0.673	/ 132.4° - 100°cos100° =	0.673 / 149.8° =	0.582 +j0.339
			Resultant =	0.0 j0.0
			=	0 /0°

Radiation Vectors as Viewed From N 147.5°E

Tower	Field	Electrical Phase + Space Phase		
1	0.538	/-123.6° + 100° cos147.5°=	0.538 /-207.9° =	-0.475 +j0.252
2	1.000	/ 0.0° + 0° =	1.000 / 0.0° =	1.000 j0
3	0.673	/ 132.4° - 100°cos147.5°=	0.673 / 216.7° =	-0.540 -j0.402
			Resultant =	-0.015 -j0.150
			=	0.151 /-95.7°

Radiation Vectors as Viewed From N 180°E

Tower	Field	Electrical Phase + Space Phase		
1	0.538	/-123.6° + 100° cos180° =	0.538 /-223.6° =	-0.390 +j0.371
2	1.000	/ 0.0° + 0° =	1.000 / 0.0° =	1.000 j0
3	0.673	/ 132.4° - 100°cos180° =	0.673 / 232.4° =	-0.411 -j0.533
			Resultant =	0.199 -j0.162
			=	0.257 /-39.1°

Figure 4.11-2. Tower radiation vectors as viewed from different azimuths.

field strength meters at each of the several monitoring points or in other critical directions. By using a two-way radio to learn the effect on field strength, cut-and-try adjustments of the phasor can yield a set of phase and ratio parameters that satisfy all conditions simultaneously.

Following readjustment, if any of the new phase and ratio parameters exceed the 3° phase and 5% ratio tolerances permitted by the FCC when applied to the parameters shown on the current station license, a partial proof-of-performance will be required before the FCC will license the new parameters. If the new parameters are within tolerance of the licensed values, a partial proof and application for changed parameters may still be desirable in order to avoid long-term operation uncomfortably close to the tolerance limits.

4.12
FM BROADCAST ANTENNAS

PETER K. ONNIGIAN, P.E.
ERIC DYE

INTRODUCTION

This chapter is for broadcast engineers, technicians and station managers who must make important decisions regarding FM transmitting antennas. To ensure the best possible signal strength in the station's service areas, the site location, antenna height, antenna type and propagation conditions must all be considered.

FM broadcasting was first authorized in the United States in 1940 by the Federal Communications Commission (FCC). The first FM station began operation in 1941. In 1945, the FM service was assigned to the 88 to 108 MHz band and divided into 100 channels, each 200 kHz wide. In 1991, there were over 5,400 commercial and 1,700 educational FM stations.

Most FM antennas are nonsymmetrical, that is they are mounted on one side of a steel supporting tower or pole. FM antennas outside the western hemisphere on the other hand are usually symmetrical, that is installed on all faces of a tower. However, both methods are capable of providing excellent omnidirectional azimuth patterns.

Antennas for FM broadcasting use horizontal polarization (H-pol), vertical polarization (V-pol), or circular polarization (CP).[1] Cross polarization is used as a means to prevent co-channel interference in some European countries but not in the western hemisphere. CP, together with its special form, elliptical polarization (E-pol), was introduced in the United States in the early 1960s as a means to provide greater signal penetration into the many different forms of FM receiving antennas, which are now found in the service area. H-pol is the standard in the United States, CP or E-pol may be used if desired. V-pol only is permitted for noncommercial FM stations seeking to limit interference to TV channel 6.

Between 1980 and 1990 there were over 300 million radios sold in the United States. Over 140 million were automobile AM/FM radios.[2] FM radio receivers use a variety of antennas including extendable monopoles (whips), dipoles and capacitive coupling to power lines and headphone leads.

Antennas for FM broadcasting must be chosen carefully in order to cover the service areas properly with adequate level and quality signals. For economic and technical reasons, the desired effective radiated power (ERP) should be produced with a balance between antenna gain and transmitter power. It is the purpose of this chapter to provide sufficient technical information for the broadcaster to achieve this.

The height of the antenna over the service area, distances to areas of population, the ERP and the economics are items that must be considered.

Antennas currently available in the United States differ considerably from those to be found in Europe. The various American types are discussed so that the engineer will be informed on the subject. Considerable advances have been made in recent years in the design and fabrication of FM antennas. These improvements provide greater penetration of signals into automobile FM radios as well as popular small FM transistor radios of all kinds.[3] The newer FM broadcasting antennas must meet the more stringent requirements for FM stereo and subcarrier broadcasting. Most FM stations in the United States are using CP antennas.

PROPAGATION

FM broadcasting has some distinct advantages over AM (medium wave) broadcast service. These advantages stem from the propagation characteristics of FM frequencies as well as the modulation system.

There is essentially no difference between day and night FM propagation conditions. FM stations have relatively uniform day and night service areas. FM propagation loss includes everything that can happen to the energy radiated from the transmitting antenna during its journey to the receiving antennas. It includes the free space path attenuation of the wave and such factors as refraction, reflection, depolarization, diffraction, absorption, scattering, Fresnel zone clearances, grazing, and Brewster angle problems.

Propagation is dependent upon all these properties out to approximately 40 miles (65 km). Some additional factors enter the picture at greater distances. Radio wave propagation is further complicated because some of these propagation variables are functions of frequency, polarization, or both, and many have location and time variations.

The technical intent of the broadcaster is to put a signal into FM receivers of sufficient strength to overcome noise and to provide at least 20 dB carrier-to-noise ratio, which will provide at least 30 dB of stereo separation. The required RF signal level varies from about 2 μV/m (microvolt per meter) for high sensitivity FM stereo tuners in the suburbs to about 500 μV/m for less sensitive transistorized portables.

Automobile receivers have wide ranged sensitivity values.

FM antenna manufacturers do not guarantee coverage. They supply antennas which meet certain radiation pattern requirements and gain. Many of the antennas assume an omnidirectional pattern. Although reasonable in free space, it is never fully achieved in practice due to sources of distortion (support structure, feed lines, etc.)

Some manufacturers in the United States provide azimuth pattern adjustment service to insure a horizontal plane pattern circularity of ± 3 dB when mounted on the side of a specific tower or pole. It must be pointed out that this radiation pattern and gain are for free-space conditions and may not relate directly to signal strengths measured at or near ground level, well away from the antenna.

Radiation pattern and propagation are two distinctly separate conditions. The pattern is the radiation which is transmitted by a given antenna in any given direction, without any propagation limitations, as measured on a good antenna test range. Propagation depends on path and environmental conditions existing between the transmitting antenna and the receivers.

The actual service area signal strength contours are based upon two probability factors. Contours are not solid signal areas. For example, the FCC signal coverage charts referred to as the F(50/50) curves, are based upon a probability of occurrence of certain voltage levels at 50% of the locations, 50% of the time. This means that at any given location, 30 ft above ground, the signal has a 50% chance to measure up to the predicted contour level. Furthermore, half the time at that location, it may reach or exceed the level predicted while at other times it may be lower in strength.

These FCC charts (FCC Rules, Section 73.333) are based upon the assumption that average propagation conditions exist. One or more of the conditions mentioned in the second paragraph under this heading may reduce the measured signal strength from the predicted values substantially.

Propagation Loss

The power radiated from a FM transmitting station is spread over a relatively large area, somewhat like an outdoor, bare light bulb on top of a tall pole. The power reaching the receiving antenna is a very small percentage of the total radiated power.

At 100 MHz and a distance of 30 miles (48 km) the figures indicate the free-space path loss to be 106 dB.[4] The formula used to compute free-space loss is:

$$FSL = 36.6 + 20 \log D(\text{miles}) + 20 \log F(\text{MHz})$$

Doubling the distance increases the space loss by 6 dB. The path loss does not attenuate the signal with distance as much as some other factors. Path loss between an earth station and a satellite is a classic text book example of a 6 dB loss every time the distance is doubled. But a typical FM station signal travels through a nearly perfect dielectric (air) and over the imperfect earth's surface (ground). Herein lies the FM radio propagation loss problem.

Refraction, diffraction, and reflection from scores of objects such as hills and buildings may occur in the propagation path between the transmitting and the receiving antennas. These, along with absorption, scattering, lack of Fresnel zone clearances, and other factors, all reduce the signal strengths.

Signal loss due to foliage has been well known to UHF TV broadcasters for many years.[5] This same condition exists to a lesser degree for FM broadcasting. Trees, shrubs, and other foliage on hills or smooth terrain affect the reflected as well as the lateral signal loss with distance. With average values of permittivity and conductivity in both foliage and ground, a loss of about 2.5 dB was found to exist in a ten-mile path, at FM frequencies.[6] The height gain factor is increased with heights above the foliage.

Considerable depolarization takes place because the transmission through or reflections from ground foliage is a diffracted field contribution.

Multipath Problems

The ideal reception condition is a strong direct single source signal. When energy from two or more paths reaches the receiver, (due to reflections) a condition called multipath reception occurs. Poor reception is experienced when there is insufficient strength difference between the direct and the reflected signals, because they can cancel each other where the geometry places them out of phase.

Nothing is more important in the way of broadcasting facilities than the location of the transmitting antenna. Great care must be exercised to find a suitable site. Poor selection of the transmitter point can result in unfavorable signal propagation and negate the entire project. One very serious result of poor site selection is multipath propagation in some directions.

As an example, the transmitter should not be located so that strong reflections take place from nearby hills or mountains. This can happen when the transmitter is placed on one side of a large city and the other side of the city has a high mountain range. Radiation into the city directly from the transmitting antenna, as well as reflections from the nearby hills and mountains will create two or more signal paths to many receivers. These reflections can be so strong that only a 10 dB difference may exist between the direct and the reflected condition which causes severe multipath problems.

A TV station at this same location would experience unusable signals due to heavy ghosting, even with directional receiving antennas, which exhibit moderate signal pickup from their back. This is illustrated in Figure 4.12-1 where a mountain range causes reflections back into a large city.

The multipath example shown in the sketch was an actual case. The site was chosen by the FM broadcaster, without proper engineering guidance, simply because the hill had a tower, building, power, and a road was in place. Later, the broadcaster learned why the original

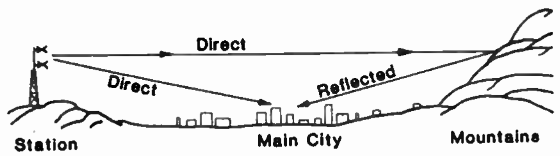


Figure 4.12-1. Example of poor station location causing severe multipath conditions due to delayed reflections from mountains on the right. Not to any scale.

owner, a television station, had abandoned the site: the TV station had failed in part due to extremely heavy ghosting into the principal city.

A much better FM transmitting site was located on the hills between the high mountain range and the city. Using a directional transmitting antenna with very little radiation towards the high mountains, reflections were satisfactorily reduced, and the FM station is now operating successfully.

Multipath reflections are easy to identify. On an automobile radio, the signal will drop out, sometimes abruptly, as the car moves. This effect may be rhythmic with distance while traveling slowly. It is sometimes called picketing as it acts like a picket fence alternately blocking and letting the signal pass. A field strength meter will usually reveal great variations of signal when moving, say, 100 ft (30 m) in a line with the transmitter. Cyclic variations over quite uniformly spaced intervals on the ground as great as 40 dB have been observed by the author.

This variation in signal levels is caused by the reflections adding and subtracting from direct and reflected signals caused by propagation problems existing in the path between transmitter and receiver. It usually has nothing to do with the qualities of the transmitting antenna. It is a function of site selection. This should not be confused with a similar effect observed near the base of the tower supporting a high-gain antenna. Nulls produced by stacking bays for gain are found near the antenna and may be filled-in if needed. (See Beam Tilt and Null Fill in this chapter.)

Ground Reflections

In the elevation plane between transmitter and receiver, nearly all FM signal coverage lies between the horizon and 10° below. Called the *grazing angle*, it lies between the horizontal plane and the earth's surface. Generally the higher the transmitting antenna above the service area, the greater this angle will be.

The angle of incidence and reflection are not the same, as shown in Figure 4.12-2. The depression angle and the grazing angle are not equal as would be the case for a flat earth. Reflections from these angles play an important part in the strength and the quality of the signal in FM broadcasting with circular polarization.

The ground, which causes reflections at these grazing angles does not treat H-pol and V-pol in the same manner. The V-pol is attenuated considerably more than the H-pol as shown in Figure 4.12-3. The phase of the V-pol changes substantially with angle, while H-pol remains nearly the same. At these useful low

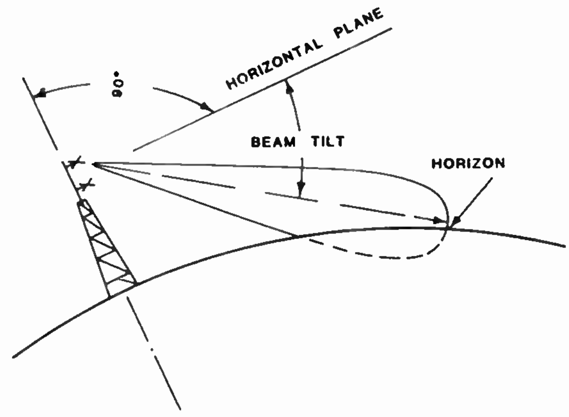


Figure 4.12-2. Beam tilt to radiate maximum ERP at the horizon. Not to any scale and exaggerated for illustration.

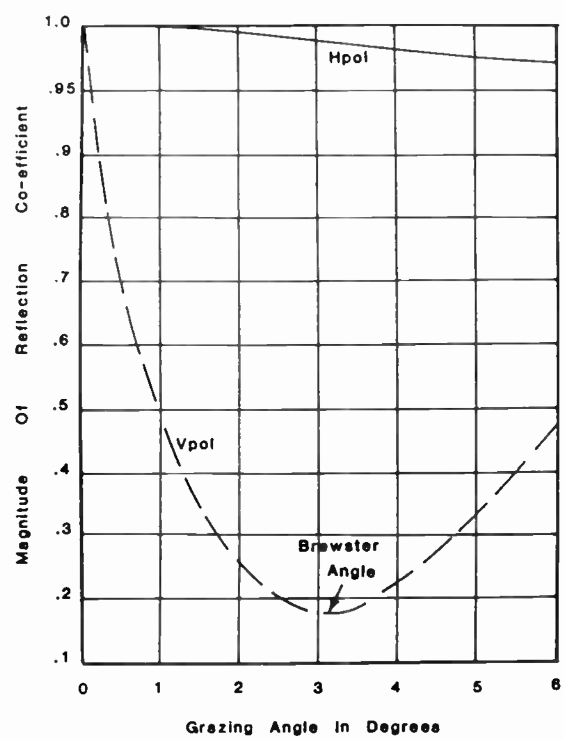


Figure 4.12-3. Magnitude of reflection coefficients showing differences for H-pol and V-pol, and the Brewster angle.

propagation angles, there is considerably less V-pol signal reflected than H-pol, when grazing takes place. Field measurements confirm this fact.[7] For this reason, it is impossible to measure accurately axial ratios in the service area. To be meaningful, the H-pol and V-pol ratios must be measured on a good antenna test range.

It is quite difficult to predict accurately the reflection coefficient (efficiency), which varies considerably as a function of polarization, frequency, grazing angle, surface roughness, soil type, moisture content, vegetation growth, weather and the season. There are complex formulas for predicting the ground conductivity at the frequency of interest. For 100 MHz, a value of 10 millimhos per meter ground conductivity is often used, with a permittivity of 25, as being about the average for the continental United States.[6]

Brewster Angle

For polarization with the electric field normal to the plane of incidence, there is no angle that will yield an equality of impedances for earth materials with different dielectric constants but like permeabilities. An incident wave with both polarizations present will have some of the one polarization component but little of the other reflected. The reflected wave at this angle is thus plane polarized with the electric field normal to the plane of incidence, and the angle is the polarizing angle.

Notice that in Figure 4.12-3 the minimum reflection coefficient occurs at a grazing angle of about 2°. Below this angle, the reflection coefficient rapidly increases to unity. The angle at which the minimum reflection coefficient occurs is called the ***Brewster*** or *polarizing angle,* after the Englishman who first discovered this phenomenon.

For ground reflections occurring near the Brewster angle, the reflection coefficient is much smaller for V-pol than the H-pol. Therefore, the reflected V-pol signal component of CP are attenuated considerably. The greatest attenuation for V-pol from ground reflection occurs at this angle.

Field measurement of V-pol signals will usually show a significant variability of H-pol to V-pol ratios due to this Brewster angle phenomenon. The Brewster angle is a function of soil conductivity and may change from place to place, as well as from season to season.[8]

It is important, then, that the antenna height above the service area results in grazing angles which are less than the Brewster angle. Otherwise the V-pol will be reduced and the radiation will be much more elliptical than circular in polarization.

Fresnel Zone Clearance

A much neglected consideration in FM transmitting antenna location and height is *Fresnel zone* radius clearance in the path to the service area. Microwave engineers always make certain that their signal paths have this important clearance.

The effect of clearance above ground or other obstacles was studied by August Jean Fresnel, a French scientist who first discovered this phenomenon in optics. Fresnel zones are circular areas surrounding the direct line-of-sight path of a radius such that the difference between the direct and the indirect path length to the zone perimeter is a multiple of half-wavelength longer than the direct path. This is illustrated in Figure 4.12-4. The zone diameter varies with frequency and path length. The greater the path length, the larger the required mid-path clearance required for full signal.

Fresnel also discovered that the entire first zone

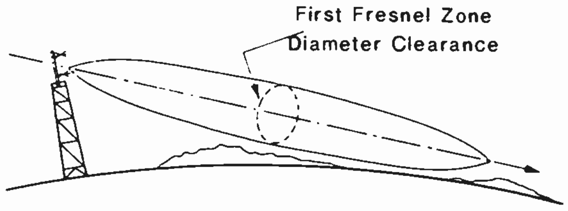


Figure 4.12-4. First Fresnel zone clearance occurs as shown above, but only six-tenths is required for full free space signal level. Not to any scale.

radius is not required for full signal strength. Six-tenths of the first zone would suffice, which is fortunate since the radius is quite large at the FM frequencies. The equation for determining the first Fresnel zone radius for 4/3rd earth curvature is:

$$R = 1140\sqrt{d/f}$$

where d is the path length in miles, f is in MHz and R is in feet for the first radius.

In Table 4.12-1 the required 0.6 first Fresnel zone radii clearances at the middle of the path are shown for 98 MHz and service areas up to 52 miles (92 km) from the transmitter. The idea is to raise the height of the transmitting antenna so that the mid-path height is as high as or higher than shown in the table. Due to the geometry of the Fresnel zone, if the terrain is relatively flat, the mid-path radius will control and be larger than that required elsewhere along the path. If the mid-path clearance is less than the values shown, the FM signal will be attenuated in accordance with the curve shown in Figure 4.12-5, presuming ideal reflection off the ground or obstructions.

The center-of-radiation heights of the antennas in Table 4.12-1 are actual and not height above average terrain (HAAT). Some of these recommended heights will reduce the allowable ERP in accordance with FCC 73.211 (b), depending on the class of station and the zone. However, it is better to have the Fresnel clearance than the maximum low height ERP values, as the higher heights will produce stronger signals.

It is a well known propagation axiom that greater heights are more useful in producing higher signal strengths far from the antenna than ERP levels, everything else being equal.

Without the first Fresnel clearance of 60% the signal level at the distant point may suffer. This reduction will follow the curve shown in Figure 4.12-5 for different values of clearance and worst case reflection conditions.

In order for the FCC prediction curves to be valid, the recommended minimum antenna heights should be achieved. These heights not only provide line-of-sight conditions to the service limits but also proper Fresnel clearances. Both conditions are required for the FCC F(50,50) curves to be valid.

The values in Table 4.12-1 are for relatively flat

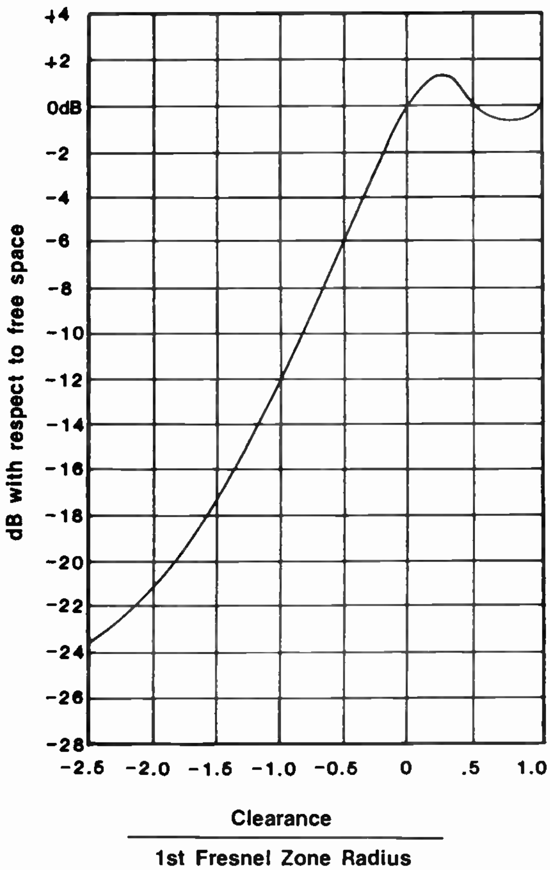


Figure 4.12-5. Attenuation of FM propagation when the path between transmitter and receiver lacks Fresnel zone clearance in the ratios shown.

terrain, but take into consideration the FCC suggested roughness factor of up to ±150 ft (50 m). Where the tower height is limited by HAAT values or other limitations, the signal strength will suffer due to those factors.

Soil Conductivity

The conductivity and permittivity of the soil, together with the vegetation on it, play a small part in the attenuation of FM signal strength. Average soil has a dielectric constant of about 15 millisiemens per meter at 100 MHz.[9]

Linear Height Gain Effect

By raising the receiving antenna above the immediate effects of the soil, the signal level will be increased. Actual field measurements have proven a 9 dB increase in signal when the dipole was raised from 3.28 ft (1 m) to a level of 30 ft (9.1 m). This is due to reflection phenomena in the foreground of the receiver, not ground conductivity.

FCC Service Contours

From the FCC coverage prediction charts, it is possible to draw contours of the various grades of service for a given ERP and antenna height above average terrain. These predictions, at 50% of the locations, 50% of the time, constitute the basis for the service contours. The city grade contour is 70 dBμ (3.16 millivolts per meter) and primary service contour is 60 dBμ (1.0 millivolts per meter).

The FCC Rules, Section 73.333 charts for these predictions have a built-in terrain roughness factor, as explained above.

GENERAL COVERAGE STANDARDS

There are certain height and power levels fixed by the FCC for various classes of stations. The United States has been divided into three different geographical areas based on population density as well as propagation refractive index levels. These ERP and height values have been set to prevent co-channel and adjacent channel interference.

Zone I, generally speaking, is the northeastern part of the United States. Zone I-A includes Puerto Rico,

Table 4.12-1
Recommended minimum antenna heights (for flat terrain and 98 HMz)

Service Area Radius Required		Fresnel Zone Six-tenths Clearance		Recommended Min. Antenna Height		Probable FCC 80–90 Class
Miles	Km	Feet	Meters	Feet	Meters	
5	8	155	47	310	95	A
7½	12	189	58	378	115	A
10	16	218	66	426	130	A
15	24	267	81	534	167	A, B, C-2
20	32	309	94	618	188	B-1, C-2
25	40	346	105	700	213	B-1
30	48	378	115	756	230	B
35	56	409	125	818	250	B
40	64	437	133	875	267	C-1
45	72	463	141	925	282	C-1
50	80	488	149	975	297	C-1
57	92	522	159	1,043	318	C, C-1

Table 4.12-2
FM station classes, zones, and ERP

FM CLASS	ZONE	MAX. ERP In kW	MAXIMUM Feet	HAAT Meters	DISTANCE TO Miles	60 dBu km
A	I, I-A	3	328	100	15	24
B	I, I-A	50	492	150	32	52
B-1	I, I-A	25	328	100	24	39
C	II	100	1,969	600	57	92
C-1	II	100	984	300	45	72
C-2	II	50	492	150	32	52

Virgin Islands, and that portion of California lying below the 40th parallel. Zone II includes Alaska, Hawaii, and the remainder of the United States not in the above two zones. This is more fully described in FCC Rules Section 73.205.

Under the FCC Rules which resulted from Docket 80–90: *Modification of FM Broadcast Station Rules to Increase the Availability of Commercial FM Broadcast Assignments*, in 1983, new ERP levels and additional classes of stations were created. The distance to the 60 dBu (1 mV/m) signal contour is the controlling factor so that the ERP based on the HAAT is adjusted to produce that level and no more at a specific distance for a particular class station.

Table 4.12-2 shows for each FM class station, the zone, the maximum ERP, the maximum HAAT, and the distance to the 60 dBu contour calculated by using the maximum ERP and HAAT, and then rounding to the nearest kilometer and mile.

Under the Rules, Class C stations are required to have at least 100 kW ERP and an antenna height of more than 984 ft (300 m) above average terrain. Class C-1 stations are now permitted a maximum of 100 kW ERP with an antenna maximum HAAT of 984 (300 m), while C-2 stations may go to 50 kW at a maximum HAAT of 492 ft (150 m). Higher HAAT may be used with reduced ERP values, in accordance with equivalent 60 dBu coverage. Class C and C-1 stations may thus share the same antenna and tower.

Stations may be upgraded using the easiest method, which is to increase existing location tower height. Such factors as local zoning laws and aircraft flight patterns may preclude this approach, however.

FM Signal Measurements

The signal strength received at 5 ft (1.5 m) above ground, which is about average for auto whip antennas, is several times lower in level than at the standard FCC measurement height of 30 ft (9.1 m). This fact should be taken into consideration when comparing low height measurements with the FCC Rules Section 73.333 prediction charts, which are based on a 30 ft receiving height, where signals are considerably stronger.[10]

Signal levels inside houses, apartments, offices, and other structures vary greatly. Levels depend on the type of building construction, but in nearly all cases will be lower than those outdoors. Reflections inside the building reduce stereo separation, and cause cross-talk problems with SCA channels. Outside FM receiving antennas generally provide good reception.

Field strength measurements should not be used to determine the transmitting antenna radiation pattern or efficiency except under carefully controlled conditions. The propagation factors discussed previously camouflage the true antenna performance. The only technically acceptable way to determine the antenna's characteristics is on an antenna test range.

See Chapter 4.15, *AM and FM Field Strength Measurements* for more information. This information may be used to determine the actual quality of service and the areas where useable signal levels in fact exist. Predicted contours may be considerably different from actual measured values.

Required Signal Strength

What is the minimum satisfactory signal strength? What is the maximum above which it is wasteful? The history of FCC proceedings provides some of the following levels:

34 dBu = 0.05 mV/m For rural areas
60 dBu = 1.00 mV/m Suburban areas
70 dBu = 3.16 mV/m Principal community
82 dBu = 12.64 mV/m Highest useful level

The first three levels were set by the FCC in the early 1950s when tube receivers and H-pol antennas were popular. Modern day transistor radios have much greater sensitivity. CP has added greater signal penetrating power than H-pol when the levels were first established.

The FCC defines two grades of signal contours on its applications. The first is based on the 70 dBu contour (3.16 mV/m) required to cover the principal community of license. The second is the 60 dBu contour (1 mV/m) which defines the primary service area.

The FCC also stated that, in rural areas, levels as low as 50 μV/m were useful. Indeed current home stereo tuners and FM auto radios operate very well with only 25 μV/m. In practice, 50 mV/m (0.05 mV/m) provides good quieting in nearly all automobile and transistor radios receiving a stereo signal from a CP station antenna. Therefore 50 mV/m should be considered the minimum useful signal level.

If the highest level of 3.16 mV/m is quadrupled, it will be 12.64 mV/m. This is a 12 dB increase, equal to increasing the FCC power level by more than 15 times. It can be safely said that this level of 12.64 mV/m is considerably more signal than necessary by any present day working FM radio. Any signal level higher than this at the receiving antenna has not proven to be of significant value.

Blanketing

Excessive RF signals can overload the front end of receivers and make satisfactory reception impossible. The FCC in Section 73.318 defines the 115 dBu (562 mV/m) level as the *blanketing contour*, and adopted the free-space prediction method to predict how far this contour extends.

New or modified FM stations have the responsibility to satisfy all complaints at no cost to the complainant, of blanketing-related interference inside this contour within one year of commencement of operations.

The distance to the 115 dBu contour is determined using the following equation:

$$\text{d (in kilometers)} = 0.394\sqrt{P}$$
$$\text{D (in miles)} = 0.245\sqrt{P}$$

where P is the maximum ERP, measured in kilowatts of the maximum radiated lobe, irrespective of vertical directivity. For directional antennas, the horizontal directivity shall be used.

ANTENNA CHARACTERISTICS

Antenna gain can be increased by adding additional radiating elements (bays) to the antenna at the cost of narrowing the radiated beam. High-gain antennas concentrate the energy into such a narrow beam that often null fill must be employed to achieve the desired signal strength within the first few miles to the tower.

Directional antennas achieve increased gain over nondirectional antennas by limiting the radiated energy in various directions. Directional antennas are useful when the tower is located near a large body of water, mountain range, or other areas where energy radiated in those directions is otherwise wasted. They are also employed to avoid interference where stations are insufficient distances apart.

Antenna gain is expressed in power ratio or in dB. For example, an antenna with a power gain of 2 is also said to have a gain of 3.0 dB.

FCC Rules, Section 73.310(a) defines antenna gain as the inverse of the square of the root mean square value of the free-space field strength produced at one mile in the horizontal plane, in millivolts per meter for 1 kW antenna input power to 137.6 mV/m. (In metric units, 1 km and 221.4 mV/m).

Notice that this gain is in reference to a horizontally polarized half-wave dipole. For a CP antenna, the gain is half for the same input power.

A two-bay H-pol antenna has a power gain of approximately two. But a two-bay CP antenna in FCC terminology has a gain of about one because the other half of the power is V-pol and is not considered in the gain calculations. Only the horizontal polarization mode is used by the Commission. The vertically polarized energy must not exceed the H-pol (except for noncommercial, educational FM facilities attempting to minimize interference to TV channel 6 reception).

The power gain of an antenna is used with the transmitter gain and transmission line loss when determining the ERP. Consider for example a 10 kW transmitter and an antenna power gain of 5. Neglecting transmission line loss, the ERP is 10 kW × 5 = 50 kW ERP. If the antenna gain were 10 and the transmitter power was 5 kW, we would have the same ERP of 50 kW. (5 kW × 10 = 50 kW ERP)

The FCC defines EPR to mean the product of the antenna input power (transmitter output power less transmission line loss) times the antenna power gain. Where circular polarization is used, the term ERP is applied separately to the H-pol and V-pol of radiation. For allocation purposes, the ERP is the H-pol component of radiation only. The V-pol component power normally must not exceed the H-pol power.

Beam Tilt

FM broadcasting antennas are normally mounted on towers which are plumb, so the peak power beam in the elevation pattern is perpendicular to the tower axis. A standard FM antenna without any beam tilt radiates more than one half of the total radiated power above the horizon. All this power is lost.

The higher the antenna is above its average terrain, the larger the predicted coverage area. Since the earth is curved, the service horizon is bent lower than a perpendicular angle from the earth's surface. Thus the strongest portion of the signal is aimed above the horizon. It also follows that the higher the antenna above the terrain, the greater the elevation angle down to the earth's horizon.

In order to strike the farthest service area from a high HAAT, the beam may need to be tilted down towards the earth. Electrical beam tilt lowers the beam angle equally in all azimuth headings and is chosen more frequently than mechanical tilting, which exhibits different effects in different directions. Choose enough tilt to position the center of the main beam on the furthest edge of the desired coverage area or just below the horizon, whichever is closer.

For low gain antennas (two to four bays), the main beam is very broad, and if the antenna HAAT is less than 500 ft there is little to be gained with beam tilt. On the other hand, beam tilt makes a large difference on high-gain antennas mounted on towers with a high HAAT.

Figure 4.12-6 shows the comparison between elevation angle path and coverage distance. It incorporates the curvature of the earth. This chart can be used to determine the optimum beam tilt. Follow the curve which is closest to actual HAAT, and mark the point where it intersects the horizon or crosses the distance of furthest desired coverage area (vertical axis). Read the beam tilt on the horizontal axis. Round this value up to the nearest ¼°.

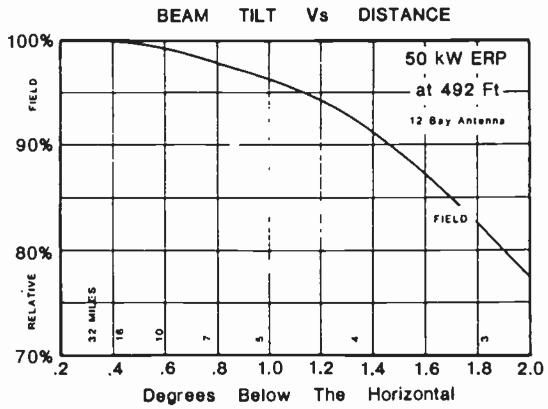


Figure 4.12-6. Twelve bay Cpol antenna with 0.3° beam tilt showing ERP distribution with coverage from 492 foot (150 m) tower over flat land. Degrees below the horizon are based on 4/3rd earths curvature. Horizon is −0.341° at 32 miles (52 km).

Consulting engineers, familiar with this problem, can easily work out the required amount of beam tilt, if it is necessary. Typical values are one-half to one degree of tilt, depending on the antenna height, distance to the far service area, and the antenna elevation pattern.

Beam tilt is usually accomplished electrically, by delaying the currents to the lower bays, and advancing the phase of the upper bay currents during the design and construction of the antenna at the factory.

Null Fill

While the beam tilt puts more signal into the far reaches of the service area, it does not solve the problem sometimes caused by high-gain antennas within several miles of the transmitter. Elevation angle nulls common to all antennas with two or more bays appear farther and farther away from the antenna as its gain is increased with more bays.

When multiple bay arrays are employed, lobes and nulls occur in the elevation pattern. As the number of bays increases, the main beam narrows and the first null radius increases. The advantage of beam tilt and null fill varies depending on factors such as tower height, site elevation, number of bays, and relative locations of communities to be served.

A simple rule-of-thumb is that null fill is beneficial when there is desired service area within the radius of the first null. The equation below gives the approximate radius of the first null for multiple bay antennas.

$$\text{Null Radius (miles)} \cong \frac{H_{(ft)}}{\text{Number of Bays}/5280\ \text{ft/mi}}$$

where H is the height (ft) of antenna above radius ring.

In most FM applications, the null is very close to the antenna, thus a small amount of null fill (5–10%) takes care of the problem. Larger amounts of null fill are unnecessary and reduce the gain of the antenna. Note that null fill has no effect on distant coverage.

VSWR Bandwidth

According to theory, the bandwidth of an FM signal is infinite if all the sidebands are taken into account. Also, at certain modulation indices, the carrier amplitude goes to zero and all the transmitted power is on frequencies (sidebands) other than the carrier frequency. Practical considerations in the transmitter and receiver circuitry make it necessary to restrict the RF bandwidth to less than infinity.

Prior to 1984 the maximum deviation for FM stations was 75 kHz, representing 100% modulation. In that year the FCC changed the maximum deviation to 82.5 kHz (110%) for those stations with 10% injection of subcarrier channels. This additional deviation requires greater antenna system bandwidth than previously needed.

System bandwidth is measured at the point in the antenna system where the transmitter is connected. This usually includes the harmonic filter, the main coaxial transmission line, and the antenna.

The significant sidebands are usually considered to be those whose amplitude exceeds one percent of the unmodulated carrier. With 110% modulation (82.5 kHz deviation) these side bands produce a bandwidth of 260 kHz.[11,12]

The VSWR bandwidth is the range over which the system under consideration has a reflection coefficient of less than five percent; a VSWR of 1.1:1.

Checking System VSWR

From time to time, the VSWR of the narrow-band antenna system should be checked and adjusted. If the exciter has thumb wheel exciter frequency adjustability in 10 kHz or 50 kHz steps, it can be used to change the frequency to check VSWR on different frequencies (with the transmitter operating at low power during the overnight experimental period). The reflectometer may be used as the indicator.

Alternately one of several methods for checking VSWR in coaxial line systems using test equipment may be used. These include a signal generator test setup, an impedance test set, or a network analyzer.

The VSWR should be measured to ensure that the reflection response is balanced to 130 kHz on each side of the carrier frequency. With transmission lines longer than 300 ft (100 m) it is suggested that the VSWR bandwidth be under 1.08:1 all the way out to ±130 kHz. The additional delay due to increasing line length becomes more of a problem, so the amplitude of the reflection must be reduced, for best operational results.

Importance of Low VSWR

The VSWR shown by the transmitter reflectometer does not increase or decrease the range of the signal. It has nothing to do with coverage. But VSWR values above 1.1:1 may decrease the final amplifier efficiency. Other definite negative effects of VSWR are increased intermodulation products and AM synchronous noise. Stereo separation is also degraded with increased VSWR.[13]

Intermodulation and SAM Distortion

Intermodulation distortion and synchronous AM (SAM) noise can be caused by narrow VSWR bandwidth in the antenna system, as well as by final amplifier circuitry all the way to and including the antenna.[14]

SAM is extremely important in FM transmitter facilities employing subcarriers. SAM is AM modulation of the carrier caused by frequency modulation of the carrier frequency in the VSWR notch. At the notch the reflected energy is the lowest. As the deviation takes place, the greater the frequency swing, the greater will be the reflections, due to the VSWR notch. With a flat VSWR curve, SAM does not take place. If the VSWR curve is skewed, SAM will occur and intermod and stereo crosstalk will increase.

Directional Antennas

The FCC sometimes requires that the azimuth radiation pattern be directionalized to reduce normally allocated ERP towards a given short-spaced station, or for other reasons. (See the FCC Rules Section 73.213, 215, and 316(b) and (c).) To conform to these specifications, most broadcasters order antennas which are pattern adjusted, measured, and certified to the Commission's requirements.

Directional antennas are licensed for peak ERP values based on the azimuth pattern. The V-pol gain may not exceed the H-pol gain in a CP directional array nor may V-pol exceed the H-pol in the protection direction (except in the case of FM protection to TV Channel 6, mentioned elsewhere). The amplitude away from the null cannot climb more than 2 dB per 10° of azimuth.

Directional antennas are usually mounted on poles although some have been tower mounted. Since the support affects the pattern, they are specified and measured with the pole or tower on which they are mounted. Most firms will make the antenna meet the specific pattern requirements.

Directionalizing is a combination of the natural pattern resulting from sidemounting and the use of parasitic elements. Using the two factors, virtually any directional pattern can be produced.

Antenna gain is calculated differently for directional antennas. The azimuth directivity increases the gain value to correspond to the pattern. If all elements/bays are the same (the typical case), pattern multiplication can be used to determine gain. For linearly polarized antennas, the gain is simply the product of the azimuth directivity, the array factor, and the efficiency factor. The array factor is referenced to an ideal dipole. For directional CP antennas, the power distribution between polarizations must be taken into account.

Antenna Gain (H-pol)

$$= g_H \times \text{array factor} \times \text{efficiency}$$

$$\text{V-pol} = g_V \times \text{array factor} \times \text{efficiency}$$

$$\text{where: } g_H = \frac{D_H \times D_V \times GA}{D_H + D_V \times GA}$$

$$g_V = g_H / GA$$

D_H and D_V are the directivities of the H-pol and V-pol azimuth patterns, while GA is the gain if the H-pol over the V-pol pattern.

Due to the gain of its azimuth patterns, directional antennas have gains which are typically 2 dB to 6 dB higher than their nondirectional counterparts of equal number of bays.

ANTENNA POLARIZATION

Radio waves are composed of electric and magnetic fields at right angles to each other and to the direction of propagation. When the electric component (E) is horizontal, the wave is said to be horizontally polarized, as shown in Figures 4.12-7(b) and 4.12-7(d). Such a wave is radiated from a horizontal dipole. References are with respect to the earth plane. If the desired electric component is vertical as in Figures 4.12-7(a) and 4.12-7(e), a vertical dipole could be used to produce the vertically polarized wave.

Circular Polarization

When the two plane waves are equal in magnitude, and if one plane wave lags or leads the other by 90 electrical degrees, the field will rotate as shown in Figure 4.12-7, at the speed of the carrier frequency and will be polarized circularly.

Only in the special case where the horizontal and vertical components are equal in strength with a 90° phase difference is the radiation said to be CP.

The direction of rotation shown by the vector arrows in Figure 4.12-7 depends on the relative phase of the two components. Thus the polarization of the wave will appear to have either clockwise or counter-clockwise rotation, as shown. The FCC has set clockwise rotation as the technical standard, in order that similar sense of rotation antennas may be used for reception in the future.

Notice that in Figure 4.12-7 the polarization rotates as the field propagates in time and space. Importantly, vertical and horizontal components are in quadrature phase. It is this rotation which enhances the signal penetrating qualities of CP, so useful in FM broadcasting.

The axial ratio as shown in Figure 4.12-8 is that between the maximum and minimum voltage component at any orientation of the reference measuring test

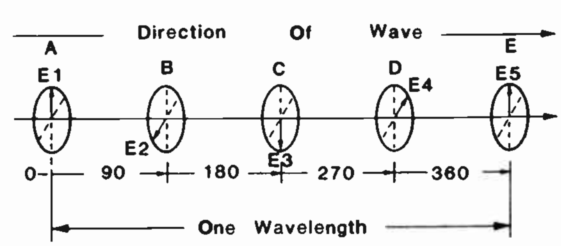


Figure 4.12-7. Circularly polarized wave propagation in one wavelength of travel, showing right hand rotation. Note vector rotation with wave travel.

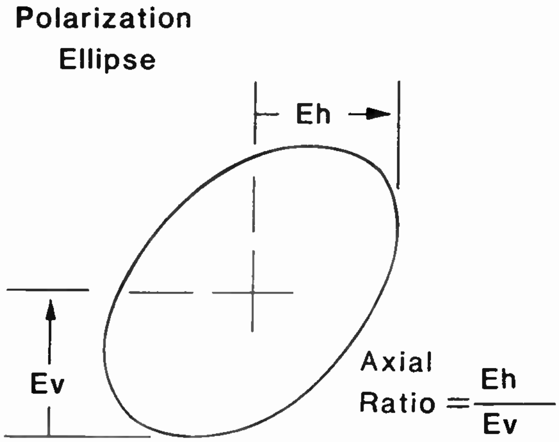


Figure 4.12-8. Axial ratio expressed in dB is the ratio of the larger polarized component divided by the smaller at any reference dipole orientation, where the maximum ratio occurs.

dipole which is placed perpendicular to the direction of propagation. An axial ratio of 1:1 (0 dB) is perfect. In practice, axial ratios of 2 dB or better are considered to be excellent and commercially available. Axial ratios over 4.9 dB (1.75 to 1 voltage ratio) are considered to be elliptically polarized, a hybrid form and not as good in signal penetrating qualities as CP.

Since receiving antennas are linearly polarized, the introduction of CP does not increase the net power received since the vertical and horizontal components never occur during the same instant. Thus, CP does not necessarily mean an increase in coverage. However, the introduction of CP eliminates the requirement of the receiving antenna to have a specific polarization. Thus, CP allows more consistent coverage within the contour. Its rotating vector can penetrate areas where linear polarization is stopped, shadowed, or canceled due to out-of-phase reflections.

MATCHING COVERAGE AND ANTENNAS

Table 4.12-3 shows the FCC predicted signal strengths for a typical Class A facility on a relatively flat plane, with the antenna center 328 ft (100 m) HAAT. A power of 3 kW is used. The first two columns show the distances, with the farthest being the horizon from this height. The third column indicates the true earth angle from the antenna to the distances shown. From the elevation information the ERP from each antenna was determined at each vertical angle. This ERP value was used to find the signal strength from the FCC F(50,50) FM prediction chart, FCC Rules Section 73.333, Figure 1.

Under the signal level in millivolts-per-meter (mV/m) column, the predicted field strengths shown in this Table are based on the above procedure. From 5 miles (8 km) to the horizon, the signal strengths are identical. This is due to the shape of the antenna elevation pattern near the maximum.

Departure occurs as the depression angle to the receiver becomes larger. Beyond 4 miles (6.4 km), the one-bay antenna and the six-bay antenna produce nearly the same signal level.

Going towards the transmitter from 4 miles (6.4 km), the field increases in favor of the one-bay antenna. In this example, the table clearly indicates that the high-power transmitter low-gain antenna does not improve the signal strength available to the receivers beyond about 4.5 miles (7.25 km). The signal level starts to increase between 4 miles and 5 miles (6.4 and 8 km). Any increase above this level is useless because full limiting has certainly taken place in even

Table 4.12-3
Transmitter power versus antenna gain
Class A 3 kW ERP — Zones 1 and 1-A
Maximum HAAT 328 Ft (100 m)

			SIGNAL LEVEL IN mV/m	
Service Miles	Distance Km	Vertical Angle	7.5 kW Transmitter 1 Bay Antenna	1 kW Transmitter 6 Bay Antenna
1	1.6	3.58	275	210
2	3.2	1.80	88	81
3	4.8	1.21	42	40
4	6.4	.92	24	22
5	8.0	.74	16	16
6	9.6	.63	11	11
7	11.3	.55	8.5	8.5
8	12.9	.49	6.2	6.2
9	14.5	.44	5.0	5.0
10	16.1	.41	3.7	3.7
12	19.3	.36	2.5	2.5
14	22.5	.33	1.8	1.8
16	25.7	.31	1.4	1.4
18	28.9	.29	1.1	1.1
20	32.2	.28	.85	.85
22	35.4	.28	.70	.70
24	38.6	.27	.55	.55
26	41.8	.27	.40	.40

the poorest FM receiver. (See Required Signal Strength in this chapter.)

In Table 4.12-3 the same signal strength of (16 mV/m) at 5 miles (8 km) comes from either transmitter-antenna combination. This is due to the fact that the ERP power at the vertical angle of −0.74° is about the same from both antennas. The ERP at 0.0° elevation pattern will of course be exactly the same for both combinations. The field does not change measurably until observation is made beyond 1.5° from the peak 100% value in a six-bay antenna.

The signal strengths in this table were based on relatively flat terrain for an antenna 328 ft (100 m) HAAT. The true earth curvature distance to the horizon is 25.56 miles (41.23 km). Therefore the outer reaches of useful signal drop off very rapidly beyond this point in the typical Class A station.

There are no nulls in a one-bay antenna pattern. In a six-bay antenna the first null occurs at about −10°, approximately 0.37 miles (0.6 km) from the tower. Antenna arrays are never perfect so the null is never zero power. With a minimum radiation of 5 W in the first null, the predicted signal would be 31 mV/m. The second null is closer to the tower and with the same 5 W ERP would be even stronger in this example. So in practice there may be no need to fill in the nulls of the six-bay antenna.

Another consideration is that the nulls may fall very close to the tower and the number of people occupying the null areas may be small. Thus problems resulting from these close-in nulls would be minor.

Typical Class B and C-2 Station Coverage

The same comparisons of transmitter-antenna combinations can be made for Class B and the new Class C-2 stations, operating with a HAAT of 492 ft (150 m) with 50 kW ERP. This is shown in Table 4.12-4. A 55 kW transmitter with a 2-bay CP antenna would provide the 50 kW ERP, with high efficiency coaxial lines. It is compared with a 10 kW transmitter feeding a 10-bay CP antenna. The terrain flatness is assumed not to exceed ±150 ft (50 m).

Table 4.12-4 indicates that the signal levels are the same from 4.6 miles (7.4 km) out to 35 miles (56 km) under similar columns as for the Class A station comparisons. The FCC uses a receiving height of 30 ft (10 m) so the horizon is a bit further away, at 31.3 miles (50.5 km).

From the transmitter out to about 2 miles (3.2 km), the signal rises much more rapidly in the 2-bay antenna than in the 10-bay, the latter being somewhat similar to a cosecant curve. There is surplus signal close in and more than is needed or can be tolerated.

Therein lies one of several problems with high transmitter power and low-gain antennas as seen in Table 4.12-4. With the 2-bay antenna there is 900 mV/m at one mile (1.6 km) and 562 mV/m as far out as 1.55 miles (2.5 km). This is above, or at, the blanketing level of 562 m/Vm. (See Blanketing in this chapter.) The high-gain antenna does not cause this type of problem under identical conditions.

The signals from both combinations are much more than necessary for present day FM receivers out to about 10 miles (16 km). There is no practical difference technically in useable signal strengths presented to receivers in the entire market area, from either antenna. There is, however, a great deal of savings in capital costs as well as operating expenses between the two combinations.

One antenna factor is not clearly indicated in Table 4.12-4. The two antennas have elevation pattern nulls. The 2-bay antenna null at −30° falls 852 ft (260 m) from the base of the tower and can be disregarded. The ten-bay antenna nulls can be filled to as little as 2½% field, which will not affect its gain. This would represent a minimum ERP at the nulls of 31 W. Although seemingly very small, it is very effective as shown in Table 4.12-5.

It is obvious that the 10-bay antenna nulls can easily

Table 4.12-4
Transmitter power versus antenna gain
Class B, C-2 50 kW EFP
Zone 1, 1-A & C-2

Service Miles	Distance Km	Vertical Angle	SIGNAL LEVEL IN mV/m: 55 kW Transmitter 2 Bay Antenna	SIGNAL LEVEL IN mV/m: 10 kW Transmitter 10 Bay Antenna
1	1.6	4.97°	900	140
1.55	2.5	3.25°	562	165
2	3.2	2.49°	310	230
3	4.8	1.67°	153	135
4	6.4	1.26°	92	88
4.6	7.4	1.15°	71	71
5	8.0	1.02°	57	57
7.5	12	.70°	22	22
10	16	.55°	13	13
15	24	.41°	6.5	6.5
20	32	.36°	3.1	3.1
25	40	.33°	1.9	1.9
30	48	.328°	1.1	1.1
35	56	.332°	.7	.7

Table 4.12-5
2½% Null fill in — 10 bay antenna

Null	Angle	ERP	Distance	Field
First	−5.75°	31 w	4,800 Ft	31 mV/m
Second	−11.50°	31 w	2,240 Ft	70 mV/m
Third	−17.25°	31 w	1,512 Ft	109 mV/m

be filled to produce signal levels in excess of those required. If the transmitter is located in a populated area, these high levels prevent the loss of stereo separation and noise in the SCA (if there are reflections from high level lobes in the built up areas). This problem is common to TV transmitters which produce ghosts from high signal level lobe areas reflecting into null areas. This problem is greatly and satisfactorily reduced with null fill as shown in Table 4.12-5.

MATCHING TRANSMITTER POWER AND ANTENNAS

Several available combinations of antenna gain and transmitter power will provide the necessary ERP. But which combination is the best? The choice is further complicated by the nature of the terrain in the service area. Is it all flat, some rolling hills, mountainous or a large valley? What are the regulatory limitations on the antenna supporting tower height?

Important considerations when choosing the transmitter power and the gain combination to produce a given ERP are as follows:

- Transmitter
- Feed system
- Antenna
- Final amplifier
- Tower
- ac power consumption

The transmitter, antenna, tower, and coaxial feed line are one-time capital costs for the station. Tube costs and commercial power use, however, are a continuing hour-by-hour cost factor. From the above it is apparent that a low-power transmitter is much more economical than a high-power transmitter. But is there a difference in signal strength?

The ERP is the product of the antenna power gain and the antenna input power. Many different combinations of power gain and input power will yield the same ERP. The azimuth pattern will be quite similar for many different antenna power gains.

The only difference in various combinations is the elevation pattern. As discussed previously, there is no momentous or important difference in serving listeners from very different transmitter/antenna ratios.

The signal strength at any given location is a direct function of the ERP from the antenna elevation pattern angle to that location, the height of the antenna, and the propagation path. The ERP at the pertinent angle is the product of the elevation pattern relative amplitude at that angle squared, times the maximum ERP. In practice there is no significant difference between a 3 kW ERP Class A station using a 7.5 kW transmitter and a 1-bay CP antenna, or, one using a 1 kW transmitter and a 6-bay CP antenna, all other factors being equal.

Normally, all the power radiated above the antenna elevation pattern to the horizon is wasted. It is the radiated power below the angle to the horizon that strikes the earth with all its FM receivers. Therefore only the radiated power towards the earth should be considered useful.

The ideal antenna system would put the same signal level from the base of the tower all the way out to the horizon. This requires an antenna whose elevation pattern is a cosecant curve, the normalized reciprocal of sine. It would be the most efficient antenna elevation pattern. Although this curve is impossible to achieve, it is approached as the antenna gain becomes greater.

ANTENNA SITE SELECTION

The transmitter location must be carefully chosen. Site economics should be secondary to the technical advantages of a particular site. Fresnel zone clearances and other factors outlined in this chapter should be considered. A site with an operating FM or VHF TV station makes an excellent source of signals to check propagation for a new station. If the existing station is FM, make certain that its antenna pattern has been optimized to provide as much circularity as possible.

A good field strength meter should be used to measure the actual signal from the existing station. Relative readings are important, not the absolute. Check for reflections as well as level changes within a short walking area of about 100 ft (30 m). Check for stereo separation. Using this information, the operation of a new station near the one being checked can be compared before moving or submitting the FCC application. The consulting engineer may find it useful to consider this information to evaluate the suitability of the new site.

High-Gain Antenna Contradictions

The many advantages of high-gain, low-power transmitter combinations to produce the ERP have been shown. Their superiority in relatively flat land applications cannot be disputed.

There is, however, the matter of unusual height over average terrain to be considered. As examples, if the transmitter is located on Mt. Wilson, in California, on a very tall building in Chicago, or New York, the elevation pattern problem can become serious. This is true particularly when there are listeners near the sites as is the case for these three locations.

Mt. Wilson which serves the greater Los Angeles metropolitan area is more than 1 mile (1.6 km) above most of its listeners. In fact coverage is required from 11 miles (17.75 km) out to the horizon which is −0.57° at 105 miles (168 km). Pasadena, the nearest city, is 13° below the horizon. A high gain antenna tilted down

0.5° would serve the far reaches well, but would not lay down a moderate signal at −13°.

Section 73.211 of the FCC's Rules limits the ERP for overheight antennas such as those on Mt. Wilson with 2,900 ft (884 m) HAAT. New stations using that height must reduce ERP in accordance with the equivalence calculation, so that the predicted signal at the 1 mV/m contour does not extend beyond 32 miles (52 km) for Class B stations.

In these situations a moderate gain antenna should be considered. From Mt. Wilson several existing four and 5-bay antennas now provide excellent service.

TV CHANNEL 6/FM ANTENNA PROBLEM

Television Channel 6 occupies the band from 82 MHz to 88 MHz. The FM broadcast band extends from 88 MHz to 108 MHz. Noncommercial educational FM stations are assigned from 88 MHz to 92 MHz. Interference can exist between the two, with the TV station viewers receiving interference from the FM stations. The FM receiver is relatively selective with a response to about 200 kHz, but the TV receiver has a bandwidth of at least 6 MHz. However, more than TV receiver selectivity is involved in this interference problem. (See FCC Rules Section 73.525.)

Three principal techniques can be employed to minimize Channel 6 interference from FM stations: (1) collocation, (2) where collocation is not feasible, location of the FM station in an area of low population density, and (3) antenna cross polarization.

Co-location

The purpose of co-location (that is, placing the FM transmitter at the Channel 6 transmitter site) is to achieve the same propagation path for both TV and FM stations, thus maintaining a nearly constant desired-to-undesired signal ratio in the service area. If possible, both antennas should be mounted on the same tower. If not, a maximum separation of 0.25 miles (400 m) between the two is still considered as co-location.

The horizontal and vertical plane radiation patterns of both antennas should be similar because the objective is to maintain a near constant desired-to-undesired signal ratio. The HAAT should be similar, thus the desirability of collocating on the same tower. The maximum ERP of the FM stations operating on this basis is in Section 73.525(d) of the FCC Rules.

Alternate Locations

The FM station may not be intended to serve the same community as the TV station, or collation may not be possible. In this event, the FM broadcaster should locate in an area of relatively low population density by imposing a limit on the population which may be included within that area where a particular undesired-to-desired protection ratio is exceeded.

Two ratios were proposed by a committee which studied this problem in 1983.[15] Their recommendation varied according to the educational station frequency separation from the channel 6 aural frequency of 87.75 MHz. In any event, the interference area should not have more than 3,000 people living in it. (See FCC Rules Section 73.525(c) and (e).)

Cross Polarization

Several organizations have made discrimination tests in the United States and in Europe with cross-polarized antennas. It has thus been well established that discrimination of 16 dB is to be expected in rural areas and 10 dB in urban areas between two stations one using the V-pol and the other using H-pol, and the receiving antenna being similarly polarized. This is sufficient in most cases to resolve the FM-Channel 6 problem.

While technically, cross polarization will help solve the problem, the Commissions Rules do not require it. This is left as an option for the FM applicant to use. Most TV channel 6 receiving antennas will remain H-pol, while automobile FM antennas will stay V-pol. So if the TV station remains H-pol, this interference problem may be cleared up if the FM station will switch to V-pol. (See FCC Rules Section 73.525(e)(4).)

Rejection Filters

The FCC believes that rejection filters installed at the TV receiver would be helpful, while others think this is not a satisfactory solution. Unfortunately, many viewers do not sufficiently understand this problem and are thus not motivated to have the necessary filter installed.

It is further complicated by the fact that many of the existing TV receivers still have balanced antenna inputs (300 Ω) and filters designed for them that do not usually provide the necessary amount of rejection. As more TV receivers with coaxial inputs are purchased by the TV viewing public, this situation could change. Coaxial (75 Ω) filters with 20 dB of attenuation of the FM signal are readily available.

COMMERCIALLY AVAILABLE ANTENNAS

There are several basic classes of antennas available for FM broadcasting. These and variations of them are made by several manufacturers in different models, gains, and input power ratings. They may be broken down into the following classes:

- Ring stub and twisted ring
- Shunt and series fed slanted dipole
- Multi-arm short helix
- Panel with crossed dipoles

These antennas have many things in common. For example, nonsymmetrical antennas are designed for sidemounting to a steel tower or pole. Radiating elements are shunted across a common rigid coax line. This has eliminated the problems associated with the older corporate feed system using semi-flexible solid dielectric low-power cables.

Shunting elements every one wavelength across a transmission line makes impedance matching simple. Bandwidth is limited by the VSWR of the individual elements and the use of an internal transformer.

With more than about 7 bays, the first three of the above antennas have a more difficult task being matched and there is undesirable beam squint, since the elevation beam angle changes with frequency deviation by the transmitter. Antennas with more than 7 bays are fed from or near the center, thus dividing the phase change in one-half and effectively eliminated the beam squint. Center feeding also simplifies the VSWR matching.

A means for tuning out reactance after the antennas have been installed on the tower is also common with all the antennas. Located at the input to the antenna, the VSWR tuner consists of adjustable location dielectric or metal slugs on the inner conductor of the main coax line. Several fixed-position variable capacitors, spaced ⅛ wavelength along the main feeder near the antenna input are also used on some sidemounted antennas to adjust the VSWR to very small values.

Another variety of antenna has curved radiating elements around a circumference whose diameter is determined by the number of element arms. Each radiator consists of two, three, or four such circular arms, depending on the model. Each element is fed through a shunt arrangement, and then shunted across the vertical rigid feed coaxial line.

Wideband panel antennas are becoming popular where high buildings, favorable mountain sites or high towers are available. Several firms make wideband panel antennas. Some have very wideband VSWR features in each radiator. Others with not so broad VSWR, use phase impedance compensation similar to the European scheme, which uses 90° phase quadrature impedance compensation.

Phase quadrature compensation makes it possible to cover the entire 88 MHz to 108 MHz band with a VSWR under 1.1:1 while maintaining excellent elevation, and azimuth patterns, together with very good axial ratios. Power ratings up to several hundred kilowatts are offered so that many FM stations can be diplexed into one such antenna.

Only the wideband community FM antenna design now uses a corporate feed system, while the others are shunt fed from a common rigid coax line. This corporate feed system, using air dielectric semi-flexible line at the lower power levels, is very successful. It splits the input power to many different dipoles at the correct amplitude and phase.

Standard Sidemount Antennas

Standard sidemount antennas come in a variety of shapes and forms. They are currently used in the majority of applications. Their chief advantages are low cost, easy installation, availability of high gains, and low tower constraints. They are available in linear polarized configurations (H-pol or V-pol) or circularly polarized (CP).

Most sidemounts are comprised of a series of radiating elements, or bays, which are fed via a rigid innerbay feed line. The most typical feed lines used are 1⅝ in. O.D. for applications with less than 10 kW antenna input power, and 3⅛ in. line for up to 40 kW. Most antenna elements come in high and low power versions. These antennas are mounted directly to the side of a tower or pole. Leg and face mounts are typical on tower structures.

Some manufacturers with test ranges offer sidemount antennas with custom directional patterns. The pattern shaping is accomplished by optimizing its mounting and adding parasitic reflectors which are on the order of ½ wavelength. Repeated range tests have shown that the sidemount antennas have largely distorted patterns due to feed lines, mounting structures, and other conductive media in the antenna aperture.

Sidemount antennas are inherently narrow banded. The bandwidth of a single element rarely exceeds 1 MHz. Although diplexing two stations on a single sidemount is occasionally done, the spread between stations must be small (a few MHz), and compensation tuning (such as long stubs must be used). In addition, none of these antennas are symmetrical, and each type has uncontrolled radiation from booms and feed lines in the aperture. This distortion deteriorates the antennas axial ratio and circularity.

Series-Fed V-Dipole Antennas

This antenna has similar bandwidth to its shunt fed counterpart, but the array is typically intentionally tuned high in frequency. The combination of this tuning scheme, and the internal protection of its feed allow this antenna to be somewhat resistant to light icing.

This model is larger in size and heavier than other types of sidemount antennas. Care should be taken in determining tower constraints. The antenna is typically field tuned for an optimized match. Careful placement of ceramic slugs can produce a good VSWR over the stations useful bandwidth.

Ring Radiators

There are several antennas that are simple adaptations of ring radiators and were designed and manufactured in the 1950s and 1960s for horizontal polarization. By adding vertical stubs to the ends of the radiator or twisting the ring, elliptical polarization (of sorts) is achieved.

Both the ring and the ring stub suffer from temperature variations which tend to change the spacing between the ring openings and thus the electrical capacitance and resonant frequency. The ring stub and the twisted ring are not really circularly polarized because the axial ratio varies considerably with azimuth. At best they may be said to be elliptically polarized.

Over the years the design has improved by adding a second horizontal ring and improving the feed. Reducing bay spacing reduces high axial side lobes. The antenna has very good circularity in free space, but like other types of sidemount antennas, it is strongly affected by its support structure and feed line.

The radiation patterns are strongly affected by the tower mounting environment. Being of relatively high Q design, they are more susceptible to detuning because of icing. Radomes and electrical deicers are available to overcome this problem. While the icing problems may be overcome, pattern optimization is not offered for these antennas.

Ring-Stub Antennas

The H-pol radiation from these antennas comes from the ring portion whose plane is parallel with the earth. There is a minor lobe from each radiator, which is strengthened with vertical stacking for additional power gain. This nadir-zenith lobe is the result of 360° stacking on the rigid coax feed line. It reduces the gain and presents a lobe at the tower base which is detrimental to low level audio equipment and personnel located in a building at the base of the tower.

In order to keep the cost down, like the twisted ring antenna, the ring-stub is manufactured in several radiator-to-radiator spacings across the FM band. This results in some minor beam tilt up or down depending on the frequency. Most higher priced slanted dipole and helix antennas are spaced exactly 360° and are usually tested to assure this spacing during production.

Shunt Fed Slanted Dipole Antennas

The slanted dipole antenna in its present configuration was developed and patented in 1970.[16] It consists of 2 half-wave dipoles bent 90°, slanted and fed inphase.

The slant angle is critical as it is the factor which determines the ratio of vertically and horizontally polarized radiated power. The phase point center is at the feed insulator on the dipole support arm as shown in Figure 4.12-9. When fed through a vertical support pole on which the antenna was mounted during initial development tests, the axial ratio varied less than 1 dB.

The commercial adaption uses a horizontal boom containing a step transformer. This boom supports two half-wave dipoles in which the included angle is 90°. The two sets of dipoles are rotated at 22.5° from the horizontal plane. Two opposite arms of the dipoles are delta matched to provide a 50 Ω impedance at the radiator input flange. All four dipole arm lengths may be adjusted to resonance by mechanical adjustment of the end fittings. Shunt feeding, when properly adjusted, provides equal currents in all four arms resulting in excellent azimuth circularity.

Short Helix Antennas

A relatively recent asymmetrical radiator is the four-arm shunt fed helix. By using four dipoles, curved so that their circumference is about one wavelength, a CP antenna is produced.[17] Each dipole is about ½ wavelength and is shunt fed. These are supported on a four arm structure, one end of which is tied to the supporting structure. The dipoles overlap so that the current flow around the circumference is circular. The four feed arms are connected in shunt and the feed impedance is quite low, but may be brought up to useful values with an internal step transformer.

Figure 4.12-9. Three bay non-symmetrical FM antenna mounted on the tower leg. The guy cables are insulated fiber glass rods near the tower legs.

The CP quality of the four-arm side-fire short helix is good. Three and two arm models are also available, but their axial ratio is not as good as the four arm. Pattern circularity is ±1 dB for the four arm, together with an axial ratio of about 3 dB.

These radiators are stacked about one wavelength apart on a rigid coax feed line to obtain the necessary power gain. Like other asymmetrical FM antennas its patterns are strongly affected by the supporting structure. See "Pattern Optimization" in this chapter for the need and methods to circularize the azimuth pattern.

Electrical deicers using the stainless steel dipole arms as one half of the heating circuit are available. Heat is created by passing a large current at low voltage through each arm from voltage dropping transformers placed at each bay level. Plastic radomes are also available to keep snow and ice off the sensitive VSWR parts of the antenna.

Twisted Ring Antennas

This type consists of one or more rings, which have been partially twisted so that the open ends of the ring are about 10 in. (25 cm) apart. One semi-circular arm of the ring is fed with a small loop or by a direct tap on that arm. A number of these rings are fed in the same manner as the ring stubs, and have the same zenith-nadir lobe problem.

The mechanical twist is not the same when viewed in all the azimuth directions. Therefore the current is not the same, with the end result that in some directions there is much more elliptical radiation than in others.

These antennas are very simple and relatively inexpensive for single frequency use, but have some serious operational limitations for CP operation. They do not have the same signal penetrating effect as the slant dipole, short helix, or the flat panel antenna type of CP antennas.

Short Helix—Multi-Arm Antennas

The number of arms may be increased to four instead of the two in the slanted dipole variety. To provide CP, the arms are curved to form a 1 wavelength circumference. These short multi-arm helices are also stacked in the conventional manner, like the others in this series for power gains as desired. This design uses 2 γ feed straps to feed all the elements in phase. This antenna is shunt fed, and is arrayed and mounted similarly to the slanted vee dipole antenna.

The azimuth pattern of all these non-symmetrical antennas is affected by the supporting steel structure. With pattern optimization, the pattern can be made quite omnidirectional. (See Pattern Optimization in this chapter.)

Series Fed Antennas

A similar arrangement of arms supported by a T arrangement may be series fed. That is, part of the outer end is insulated from the rest of the dipole and fed across the insulated break. To allow for adequate power handling capacity and to increase the VSWR bandwidth, 3 in. (75 mm) diameter tubing is used.

The antenna has an VSWR bandwidth of about 1%, so it makes an excellent single channel FM antenna. The antennas are usually mounted on the side of a supporting tower or pole, and stacked vertically to achieve required power gain.

This antenna has greater wind loading than the shunt fed version due to its larger element diameters necessary to achieve useable VSWR bandwidth. It presents considerable large amounts of ac power for electrical deicing. Plastic radomes also present additional wind loading.

Flat Panel Antennas

Panel antennas for CP FM broadcasting are relatively new in the Western Hemisphere, although H-pol and V-pol have been used in Europe since the mid-1950s. This antenna was developed in Europe to provide a wide bandwidth for several collocated government stations without the need to change antennas when a new channel was added or the operating channels were changed from time to time.

Panels are from 7 to 8 ft (2,100 to 2,450 mm) square in the flat configuration. In the cavity style they are about 8 ft (2450 mm) in diameter and about 3 ft (1,000 mm) deep.

A heavy metal frame is often used over which large diameter wire mesh has been welded. The wire mesh screen openings vary from 4 to 12 in. (100 mm to 300 mm). Electrically they are considered nearly solid metal. These openings produce relatively low wind loads. The entire flat frame or cavity is strong enough to support a man on its mesh openings. Some manufacturers hot dip galvanize their steel after fabrication; others use stainless steel construction.

For FM use, two crossed dipoles are used as the illuminating source for each panel or cavity. Each dipole is fed in phase quadrature, that is one dipole receives its peak current 90° after the other, to produce CP. A typical set of electrical and mechanical specifications for a CP 8-bay cavity community antenna is shown in Table 4.12-6.

Flat panel antennas are typically sidemounted on large face size towers. The screen panels greatly reduce interaction and distortion between the antenna and tower. The panels are directional, thereby requiring 3 or 4 panels to be mounted around the tower to achieve acceptable azimuthal circularity.

These antennas are usually branch fed, and often the array's top and bottom halves are fed separately. This allows operation of either half of the antenna separately when it is necessary for maintenance or repairs. Circular polarization is achieved on each panel by feeding two perpendicular dipoles 90° out-of-phase. This phase offset helps this antenna achieve usable bandwidths on the order of 10 MHz.

By pulling the dipole back on its feed support arms, the arrowhead shaped dipoles control both V-pol and H-pol azimuth patterns. Rotating the dipoles 45° with the earth-ground reference improves the polarization ratios even further.

Round dipoles made of tubing as large as 6⅛ in. (155 mm) in diameter are used along with a single line quadrature feed. This combined arrangement makes an excellent wideband CP panel to cover the entire FM band. Power splitters, dividers, and cables, along with a number of these panels, completes the antenna.

Even on large face towers, circularity in the H-pol can be quite good, on the order of ±2 dB. On standard

Table 4.12-6

Typical measured community antenna performance

Operational frequency range	88 to 108 MHz
Safe RMS input power rating	200 kW
Power gain ratio, each polarization	4.4 (6.43 dB)
Maximum VSWR any frequency between 88–108 MHz	1.1:1
Elevation pattern beam tilt	−0.5°
Polarization	Right hand circular
Axial ratio	Better than 2 dB
Azimuth circularity Vpol or Hpol	Better than ±2 dB
Antenna dead weight, less than	7,000 Lbs (3,183 kgs)
Active wind load, RS-222-C 50/53 PSF	8,000 Lbs (3,636 kgs)
Antenna input flanges, two, size	6-1/8 inch
Number of bays (stacks)	Eight
Radiator type	Circularly Polarized Cavity

configurations, the V-pol pattern is quite different. As a result, the axial ratio of this antenna ranges from good at some azimuth headings to rather poor at others. This is because the azimuth pattern of a H-pol dipole is like a figure eight, or cosine function, while the pattern for a V-pol dipole is not directional in its azimuth plane. Therefore, each polarization will react quite differently when mounted in front of a panel.

Dipoles on these panels are often mounted at 45° referenced to the ground. This has no effect on the axial ratio or pattern performance, it instead is done for tuning considerations to compensate for mutual coupling.

Variations to this design have reduced the differences between the patterns of the polarizations. These techniques are effective for applications requiring only a few MHz of bandwidth. One method optimizes the angle of the dipole bend as well as its distance to the panel. This design requires three panels to be mounted around a tower. Axial ratio and pattern circularity are improved at the cost of system bandwidth.

Another method uses four dipoles forming a square shape in front of the panel. By adjusting the spacing between the dipoles, the beamwidth of a panel can be controlled. Over a small bandwidth, the pattern performance is greatly improved. It is necessary to mount four panels around a tower for a circular pattern. A large amount of panel interaction and leakage are severe design limitations.

For projects that require wider bandwidth, skew mounting is often used. This physical configuration allows the panels to be fed in mod 1 (0°, 90°, 180°, 270° phase for 4 around) which can increase the bandwidth of the system at the input. Although skew mounting deteriorates pattern performance, the increase in bandwidth extends its applications.

Cavity Backed Panel Antennas

The use of a cavity screen instead of a flat panel has greatly improved axial ratios. The cavity acts as a resonator with little leakage towards the tower. The shape of the azimuth pattern in each plane becomes both controllable and symmetrical. System bandwidth is improved over the flat screen design. By adjusting the diameter of the cavity structure, beam widths can be altered to meet specific requirements. Mounting three cavities around a tower gives good pattern circularity. Axial ratios usually range from good to excellent.

The cavity antenna uses the reflective properties of the flat screen panel. In the cavity however, the illuminating dipoles are flat instead of round and all four arms are parallel to the plane of the cavity. Like the flat panel with its round dipole supporting balun, the cavity also holds its flat dipoles with a double coaxial balun.

The dipoles in the cavity get their wide VSWR bandwidth through the sleeve dipole principle.[18] Capacity is provided by a metallic ring close to all four dipole arms placed between them and the back of the cavity. Circulating surface currents flow on the dipole arms, which result in evenly radiated patterns in all polarization planes. The bandwidth of a single cavity can cover the full 20 MHz band with a VSWR better than 1.1:1. Therefore it is not necessary to skew mount these antennas for bandwidth considerations.

This antenna has the advantage over some other designs of greater VSWR bandwidth. It is considered closer to state-of-the-art due to better elevation and azimuth pattern control of both planes of polarizations by the shape of the cavity.

Cavities and flat panels can be modeled using a computer. Factors such as tower size and orientation, as well as the phase and skew of the elements, can be modeled to determine optimum mounting and feeding. This is useful in projects which require a directional pattern. A station should take pattern constraints and desires to an antenna manufacturer to find out what is feasible.

Crossed Dipole Theory

Common to the flat panel and the cavity is the operation of the dipoles which generate CP. The dipoles are fed currents in phase quadrature, through a coaxially balanced balun, which provides equal currents to all four arms of the two dipoles. They excite the entire cavity or flat panel with a rotating RF field in a plane parallel to the dipoles. The RF field is thus CP and may be ideally represented by a rotating vector of constant magnitude revolving one revolution per wavelength of propagation distance. It is right hand polarized as the field rotation is clockwise as viewed from behind the screen, looking toward the direction of propagation, if the phasing between the two crossed dipoles is properly made.

Radiation patterns, associated beamwidth, and directivity are determined to a large extent by the size of the cavity or flat panel. The geometry of the dipole has less effect than the reflector size. The size and shape of the dipole controls the antenna impedance and the VSWR. The screen panel, be it flat or a cavity, fulfills the following five important electrical functions:

- Isolates the radiating elements from the tower or the mounting structure, and reduces mutual coupling
- Provides sharper beamwidth and more gain than achievable with the dipoles alone
- Furnishes pattern control so that the beamwidth is nearly equal for both horizontal and vertical plane polarization
- With an effective balun feed system, the crossed dipole radiated pattern phase is very uniform as the amplitude changes normally with azimuth
- Computer aided designs are easily achieved in production for various width towers because the pattern is simply pure electrical geometry.

Antenna Element Spacing and Downward Radiation

Most FM antennas have elements that are spaced one wavelength apart (9 to 11 ft) for reasons such as gain considerations, mutual coupling effects and ease of feed design. There are cases, however, which require a different element spacing. High levels of downward radiation is the most common reason, although considerations such as aperture constraints and beam shaping also utilize nonwavelength spacing.

Radio frequency radiation (RFR) safety levels must be considered in nearly all site locations. Power radiated in the lowest sidelobe can cause a variety of problems, of which human exposure levels are most critical. See Chapter 9.4, *Non-Ionizing Radiation,* for further information.

When antenna elements are arrayed, the resulting elevation pattern contains lobes and nulls. The furthest sidelobe from the horizon typically peaks between 70° and 90° below the horizon for full wave spacing. This lobe occurs since the physical path results in no phase cancellation in that direction, and thus each elements downward radiation is additive. Shortening the spacing changes the difference in the elements physical path length, and results in some phase cancellation.

Half wave spacing, for example, greatly suppresses the levels of the side lobes, while increasing the width of the main beam. Despite the lack of power in the side lobes, the extra width of the main beam causes the pattern to be less directive. Thus, this exchange results in an overall gain reduction on the order of a third. Spacing the elements 0.8 wavelengths apart improves sidelobe suppression and the gain of the antenna is not greatly affected.

Short spaced antennas are fed either by a shunt or branch feed system. For half-wave spaced antennas, a shunt line delivers each element 180° out-of-phase with the next element. This phase distribution problem can be overcome by flipping every other element upside down, thus inducing a 180° phase shift in the feed. Other spacings, such as 0.8 wavelengths, require a branch feed, which can deliver equally phased signals to each element independent of spacing.

Problems With Sidemounted Antennas

Single station FM antennas are typically sidemounted on a pole or tower. Unlike panel antennas, the support structure greatly effects or distorts the radiation pattern. The resultant pattern may have large peaks and nulls that can result in coverage and reception problems.

In addition, the V-pol and H-pol patterns react quite differently to these distortions. Due to the geometric complexities of the CP radiating elements and tower structure, computer modeling exists to accurately predict pattern effects is a difficult proposition. Therefore, the use of test ranges are required to determine how an antenna behaves when mounted on a tower section similar to the one on which the antenna will eventually be used.

For nondirectional stations, a test range can determine the proper mounting of an antenna to achieve an acceptable circularity. Depending on the tower size, the depths of nulls can be greater than 10 dB. An optimized mounting configuration can make the nulls less significant and oriented in areas where service to the primary coverage area is not hurt. Parasitic reflectors are often used to improve the circularity of nondirectional antennas.

When the top spot on a tower or structure is available, pole mounting is often preferred. A pole provides a stable and symmetrical support structure which has low interaction with the horizontally polarized component. In combination with the feed line, the pole typically induces a null in the V-pol pattern, directly opposite of the elements. Proper orientation of the element can reduce the effects of pattern distortion.

Mounting an antenna on the side of a tower can produce unpredictable results. The position of the peaks and nulls vary greatly from the orientation of the elements. As the face size of the tower increases, the distorting effects magnify. To compensate, many stations use smaller sections of tower at the top, where they plan to install the FM antenna. Eighteen-inch and smaller face towers tend to produce good results. With careful planning, the use of a 24 in. or larger face tower can also be successful.

Note that the patterns of each polarization react differently, and thus axial ratios can be quite poor.

PATTERN OPTIMIZATION

Single-station FM antennas are usually sidemounted on a pole or tower. This is economical and it frees the tower top for other possible uses. Unfortunately the pole or tower tends to distort the radiation pattern, seriously affecting station coverage in some directions.[19]

This problem can be serious if the FM antenna has been randomly attached to a support tower. FM antenna makers do not manufacture and sell towers. A few have made supporting poles on which the FM antenna has been affixed, adjusted and pattern tested. TV antenna makers, on the other hand, usually build the antenna as a complete self-supporting structure to be mounted on top of a support. They are usually not faced with this sidemounting problem. The logical but more expensive solution would be to make the FM antenna a self-supporting structure just like TV antennas.

Improper FM antenna sidemount installation on a tower can cause serious pattern problems. Measured patterns have indicated that, in some cases, the maximum radiation can actually be in the opposite direction from the desired direction.[20]

Need To Optimize

It is not wise to gamble with an FM station's antenna coverage. Nulls may be toward important service areas. Nulls as low as 1% of the RMS power have been measured with towers varying in width from 18 to 120 in. (0.5 m to 3 m).

Another problem is that with nulls come lobes. Lobes as great as 9.8 dB over RMS have been found. When used without pattern optimization, this lobe would produce an ERP in a given direction nearly ten times the FCC licensed value.

Translating this to a 50 kW ERP station there would be radiation in some directions of only 0.5 kW and others with 477 kW. This is a maximum to minimum ratio of 29.8 dB, and clearly not acceptable.

With CP came additional problems as the H-pol and the V-pol ratios are not always the same and vary moderately in any given azimuth. This ratio can be as great as 15 dB and must also be addressed in order to resolve the horizontal plane circularity problem. The axial ratio could be degraded causing the V-pol radiation in certain directions to be much stronger than the H-pol. This violates FCC's requirement that with CP, the V-pol must not be stronger than the H-pol component.

Section 73.316 of the FCC Rules covers FM antennas but does not specifically address this problem of azimuth circularity. In fact the FCC assumes that FM nondirectional broadcast antennas have perfectly circular horizontal radiation patterns.[21] In actual practice, they seldom do.

In order to produce a horizontal plane pattern which even approaches a circle requires considerable work by the firm making the antenna. Since it is nearly impossible to produce a circular pattern with a non-symmetrical sidemounted antenna, the term and technique, *optimization* (to do the best possible) is in common usage now.

Theory of Optimization

Figure 4.12-10 indicates how energy from the horizontal loop representing H-pol is intercepted by a pole and reradiated. Similarly, in Figure 4.12-11 energy from the vertical dipole is intercepted and reradiated as V-pol.

In the first case, the pole diameter is small in wavelength in the direction of the electric field, thus, scattering is minimal and not much H-pol radiation can be expected. However when the V-pol dipole excites the pole, a large amount of energy is intercepted (Is) and

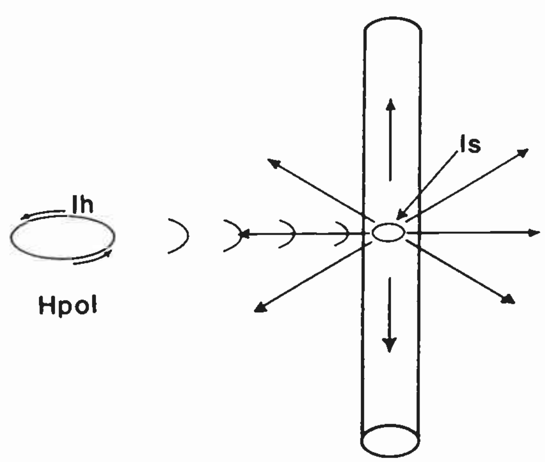


Figure 4.12-10. The effects of supporting steel towers or poles on one side of a non-symmetrical antenna are shown as horizontally polarized currents Ih flow on supporting members as Is and reradiate in all directions.

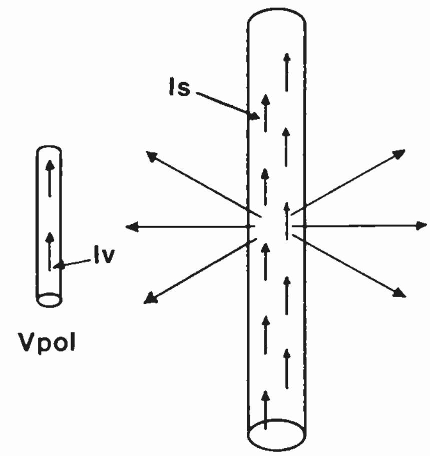


Figure 4.12-11. Vertically polarized current Iv flow over the supporting structure members as Is and reradiate in all directions.

reradiated because the large dimension of the pole (length) is parallel to the electric field. A similar effect is produced by the vertical transmission line which is common to the antenna itself, (IV in Figure 4.12-11). The result is appreciable distortion of the vertically polarized azimuth pattern.

Figure 4.12-12 shows the resulting H-pol and the V-pol patterns. The pole and/or vertical coaxial line have transformed the V-pol pattern from circular to a cardioid, while the H-pol pattern remains essentially omnidirectional. The null of the cardioid can be more than 7 dB down from the RMS value. This phenomenon is well known, and as a compromise, broadcasters generally install the antenna on the side of the tower support structure facing the main service area. There are many exceptions to this, as some measured patterns

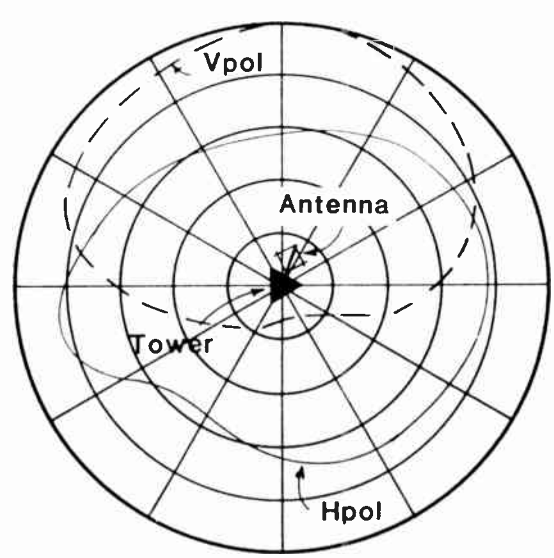


Figure 4.12-12. Combined results of H-pol and V-pol reradiation distort both patterns, producing more than 17 dB variation. This poor Cpol pattern should be pattern optimized to smooth out the azimuth pattern.

on triangular towers of standard construction have shown.

In Figure 4.12-12 the V-pol is much stronger than the H-pol in the favored direction. In the opposite direction, there is little V-pol. The power ratio of H-pol to V-pol is 16 times. This makes for a very poor CP antenna in some directions.

Tower and poles under about 2 ft (0.6 m) in cross section will exhibit the same effects on the antenna patterns, as in Figure 4.12-12. Towers greater than this size obviously will increase the complexity of scattering effects. Three or four tower legs, the horizontal and diagonal cross members, transmission lines, ladders, tower lighting, and deicer conduits, all will be excited by the vertical and horizontal currents from the radiators. All these surfaces will reradiate and affect the horizontal plan patterns.

In contrast to the simplicity of the antenna on the side of a pole, the tower supported antenna may be mounted on the face, or on a corner, at or between horizontal cross members, or tilted at various angles compared to the tower—all multiplying the complex factors affecting the patterns.

Optimization Methods

The most popular technique to achieve the desired pattern is through the use of Yagi antenna principles, wherein parasitic elements are placed in the field of the radiator to modify its radiation pattern. For example, a shortened dipole (director) placed in close proximity to a radiator reinforces radiation in the forward direction, and suppresses the signal in the opposite direction. If the parasitic element is longer than the radiator (reflector), the effect is reversed. The signal is suppressed on the side of the parasitic element and reinforced in the direction of the radiator.

Similarly, parasitic elements can be used with FM antennas mounted on the sides of towers or poles to produce pattern changes. As discussed here, both directors and reflectors may be used. Both are frequency sensitive. The effects of the supporting structure are also frequency sensitive.

Therefore, an arrangement of parasitic elements for a given FM frequency will not necessarily be the same for another, nor will the pattern be the same for a given arrangement, if it is moved up or down the tower by as much as 1.5 ft (0.5 m).

The resulting patterns cannot be predicted. There are many factors which affect the horizontal plan pattern. Only by actual antenna pattern range testing can the patterns be adjusted and properly measured. Therefore the cost for doing this is high, since it is time consuming, and requires qualified antenna technicians and the use of an antenna range. In addition, the final parasitic arrangement must be permanently fabricated and installed. However, the results are well worth it.

Pattern Service

There are two basic types of pattern service furnished by several antenna manufacturers in the United States. FM antennas may be adjusted for the best omnidirectional pattern possible or they may be adjusted to proven minimum ERP values in particular azimuth directions. The minimum required values, plotted on a polar chart by the broadcasters may be combined with the tower orientation. Using the customer's make and model tower, two or more bays of the antenna are fabricated, installed on a section of the tower, and put on the test range. Adjustments are made such as leg chosen, distance from the leg, and the orientation of the antenna with respect to that leg. Parasitic elements are then used to further improve and shape the pattern, so that the minimum ERP values will be achieved in the customers service area as given with the order on the polar plot.

For example, a Class A station may wish that a minimum of 3 kW be radiated in a pie from 90° to 120° and the remainder of the azimuth be no less than 1.5 kW. This would then require the antenna technicians to achieve a pattern without any field voltages less than 70% and that the vectors between 90° and 120° to be 100%. This sort of work has been done by many antenna manufacturers. See Figures 4.12-13A and 4.12-13B for a typical before and after optimized pattern.

Various methods have been used to optimize FM antenna azimuth patterns. Some firms use models at twice the operating frequency. Others use theoretical methods, backed up by experimental proof. The final optimized antenna is match marked on the tower sections so that it will be assembled exactly as it was made and tested at the fabricator's plant, and tested on the antenna range.

A complete set of installation prints must be provided so that the antenna is assembled exactly as tested, with all the correct locations, and angles of all the parasitic elements.

WIDEBAND COMMUNITY AND MULTI-USER ANTENNAS

The most significant developments in FM antennas have come in broadband designs. Flat panels, cavities, helicals, and a sidemount design have given broadcasters a range of choices to fit their application. Factors such as pattern requirements, tower constraints, and system budget can narrow the options.

In order for an antenna to be useful throughout the 20 MHz FM band, its operation must be the same on any frequency. The VSWR on 89 MHz for example, must be just as good as on 108 MHz. The CP azimuth pattern should remain the same on one end of the band as the other, as must the axial ratio. This is a much more rigorous requirement than placed on the single-channel slanted dipole or the ring stub.

In the wideband antennas, several factors go together in order to meet these severe requirements:

- Basic wideband dipole radiators
- Screen-panel pattern control
- Quadrature phase distribution

By using these three principle parameters in a wideband antenna the radiation pattern, VSWR, and the

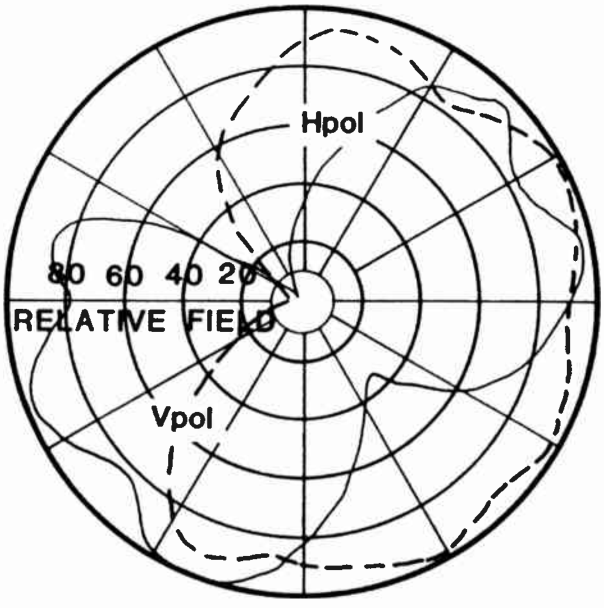


Figure 4.12-13(a). Measured non-symmetrical C-pol pattern of tower side mounted antenna. V-pol variation is ±15 dB, while H-pol is ±12 dB. The axial ratio was 24 dB. Antenna patterns are very poor.

gain can be nearly the same on any channel within the FM band.

Several methods are used to make the VSWR of the crossed dipoles as good as possible. The dipoles are usually fed with a folded balun or the split-tube type balun.[22] This improves the impedance match, phase as well as amplitude linearity of the resulting azimuth pattern.

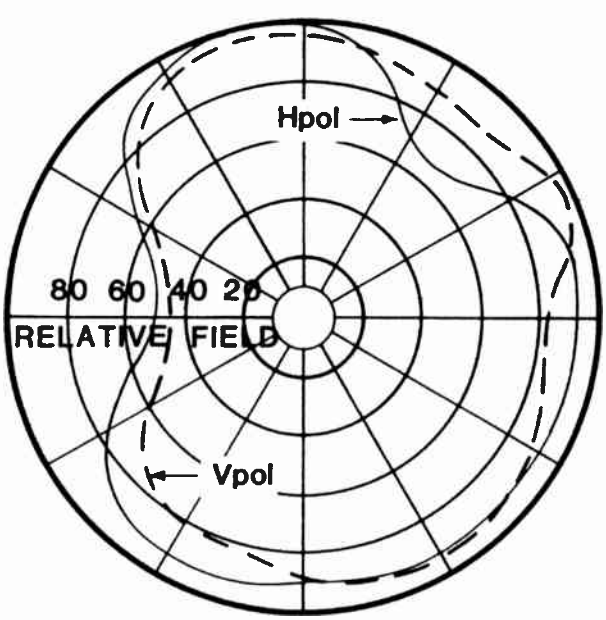


Figure 4.12-13(b). Same antenna as 22A but much improved after pattern optimization. H-pol variation is ±3 dB and V-pol is ±3.4 dB. Axial ratio is quite acceptable at 2.9 dB.

The length to diameter (or width) ratio is usually about three. This not only reduces the Q but also increases the voltage flash over levels. The low Q also increases the bandwidth by decreasing the rate of reactance change with frequency.

A natural factor aiding the VSWR problem is the fact that in order to obtain CP from two crossed dipoles, there must be a phase quadrature of the two currents feeding the crossed dipole. The two reflections, as a result of VSWR, return back to the phasing device 180° out-of-phase with each other. Being the same amplitude, they cancel.

All of the above factors, plus two or three more levels of quadrature reflection cancellations, bring the overall system input at the antenna to under 1.08:1 across the band. This cancellation technique eliminates the need for electrical deicers or plastic radomes, as the VSWR is not affected by moderate ice coatings. However radomes may be necessary with flat panels, to physically protect some radiators from falling ice.

These and other factors all contribute to make the panel-cavity antenna the best possible for either single channel or community antenna use.

Top Mount Antennas

There are a few broadband designs which require top mounting on a pole. The first design incorporates a series of dipoles which are mounted on a pole. Each dipole is branch fed individually. By rotating the mounting scheme, circular patterns can be achieved. The rather small radius between dipoles keeps nulls from crossover points to an acceptable level. Large bandwidths and good patterns have been measured on this system. The considerable amount of feed cables that run through the aperture, however, can cause problems if not properly RF grounded or shielded. Line damage, failure, and pattern distortion are possible results.

Spiral (or normal mode helical) antennas are now being used for FM applications. This traveling-wave-type antenna has bandwidth on the order of 15% to 20%. Fed from the top, a series of wires are wrapped around the conductive pole over an aperture of 2 to 3 wavelengths. The pitch angle of the wire wrap controls the pattern characteristics. Pattern circularities of ±1 over a 20 MHz band have been measured. Axial ratios better than 2 dB are considered typical. The feed system is enclosed by stainless steel feed cans which eliminates pattern distortion by the feed system, a common fault of other top mounts. In icing regions, heating elements inside the radiating wires stabilize its performance.

Multi-Channel Sidemount

Another broadband FM broadcasting antenna is the sidemount. This design incorporates dual baluns, which insure a balanced and symmetric current excitation. This balance eliminates spurious radiation from currents on booms, feed lines, and other nearby conductors. The feed arms are of the skewed vee dipole configuration. Excellent circularities and axial ratios have been measured in free space. Interaction with feed lines and support structure still distort the pattern, though to a lesser extent.

A single FM element has been measured to have a bandwidth of 5.5 MHz with a VSWR of 1.1:1. This bandwidth does not limit the spread of two combined stations. By centering the band edges, the antenna can tune two stations over 8 MHz apart with a VSWR of 1.1:1 over each channel. This comes at a cost of poor performance at center frequencies, thus three stations over 8 MHz is not achievable. As a sidemount, this antenna has substantially reduced costs and tower constraints over a panel type.

Community Antenna Economics

The community antenna fits best in multiple station service. This allows sharing costs so all parties benefit from a superior antenna that each station independently could not economically justify.[23]

If enough planning is done in advance, it may be possible to install all the FM stations of one community on one tower, at considerable savings to all users. Some preclusions are lack of adequate mileage separations, the existence of excellent facilities, and FAA tower height limitations.

The break-even point appears to be with four stations. When five stations are involved, there is perhaps a 20% reduction in cost to each of them when compared to the costs of putting up their own individual single channel antenna and tower at the same height.

One of the first large-scale community antenna projects in the U.S. was the Senior Road Tower.

In 1984, a group of Houston, Texas broadcasters formed the Senior Road Tower Group, and installed a 2,049 ft (625 m) tower supporting a 12-bay community FM antenna system, with its HAAT at 2,000 ft (610 m).[24] This height permitted the maximum service allowed.

Two runs of 8³⁄₁₆ in. (208 mm) diameter coaxial lines are used to feed the antenna in such a manner that power in both the lines causes right hand CP from the antenna for nine FM stations.

The nine stations use one diplexer each, all of which are housed in one 2,400 square foot (223 square meters) room. The 10 port modular diplexer has a total power handling capability of 350 kW. The insertion loss for each station is 0.80 dB (17% loss). The isolation between the various transmitters meets FCC spurious emission (intermodulation) requirements.

All the diplexers are monitored at a central operating rack which indicates each diplexer's forward, reflected, and rejected power. This permits trouble shooting in an orderly and rapid manner. Electrically operated coaxial switching permits each station to be connected to the dummy load for individual testing. Air conditioning and chilled water are used to remove heat produced during operation.

This community FM antenna project demonstrates the technical feasibility for a large system with nine FM users. The technology is available, the difficulty is in bringing broadcasters together to start, plan, fund and complete such a large undertaking.

Technical Advantages

Besides the financial advantages cited under the economics heading, there may be the competitive advantage of protecting the channel classification and using the same height antenna as the competitor. Other advantages include its emergency upper-lower half feature for transmission line or antenna half backup. The flat VSWR curve is highly useful for SCA and stereo operation. Stations sharing this type antenna will all experience less intermodulation interference than if they had separate antennas, but closely placed.

STATION DIPLEXING

Diplexers are passive devices used to combine the power of two or more stations and feed the combined power to a common transmission line and/or a common transmitting antenna. This system of utilizing one well-sited, high-quality antenna has become popular, convenient and economical.

Wideband panel antennas, although expensive for use by one station, are cost-effective for two or more stations. These antennas maintain their omnidirectional horizontal plane patterns and VSWR throughout the FM broadcasting band from 88 to 108 MHz. Thus, they make the ideal antenna for multi-station diplexing.

Diplexer Economics

Diplexers permit several stations to be combined into a single master antenna. New channels at any power level may be combined in any order by connecting another diplexer in tandem to the previous unit. The first channel may be fed into a wideband antenna without a diplexer. When two stations are to be combined, the first diplexer is put into service. When the third station is added to the system, a second diplexer is required and so forth. Figure 4.12-14 shows this arrangement.

Each diplexer contains a pass through broadband input port and an injected frequency port. Nearly all uncombined diplexer power and other undesired products are absorbed by a dummy load connected to the fourth port. The third port goes to the antenna or to the next diplexer's broadband input port.

The combining of two different stations must be done without degradation to station. The important factors are amplitude, group delay, VSWR bandwidth, isolation, and insertion loss. Insertion loss is a continuing cost item as it consumes RF power that is generated by the transmitter at considerable expense. Table 4.12-7 shows the measured performance of a typical three-diplexer installation.

Constant Impedance Type Diplexers

The constant impedance type takes its name from an operating characteristic of one of its components,

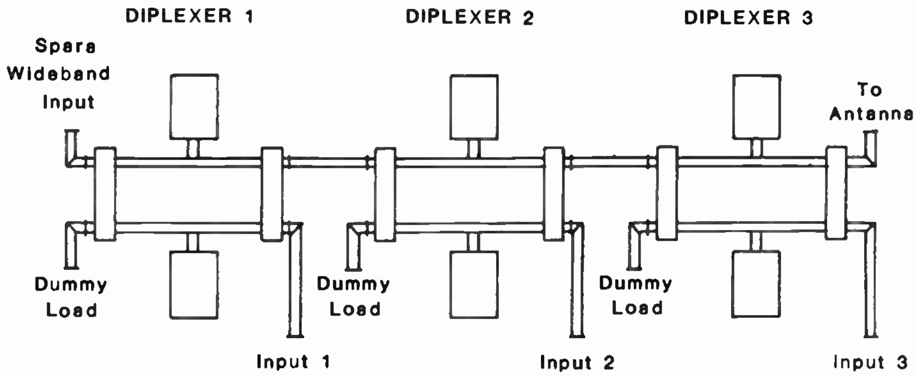


Figure 4.12-14. Three diplexers connected to provide combining facilities for three different stations for one community antenna. Note spare emergency wideband input port.

Table 4.12-7
Typical measured diplexer performance
(Three channels into one antenna)

Channel frequency	93.5 MHz	99.3 MHz	100.5 MHz
Input VSWR, plus minus 250 kHz	1.1:1	1.04:1	1.04:1
Isolation: From 93.5	—	−80 dB	−80 dB
From 99.3	−59 dB	–	−50 dB
From 100.5	−80 dB	−50 dB	–
From broad band port	−56 dB	−44 dB	−39 dB
Amplitude Response:			
Carrier plus minus 250 kHz	0.15 dB	0.25 dB	0.35 dB
Group Delay, in nanoseconds:			
Carrier plus minus 250 kHz	25	27	38
Insertion Loss, carrier	0.23 dB	0.24 dB	0.23 dB

the constant impedance of the broadband 3 dB hybrid. All of this type use two hybrids and cavities. Because of the use of hybrids, the constant impedance hybrid diplexer is sometimes called a *hybrid combiner*.

An efficient diplexer is not a complicated device, as it consists of two basic components, the two hybrids and two cavities. Coax is used to interconnect the components. Hybrids and cavities either have coaxial flange connections or are directly coupled. Some units are pressurized to keep moisture from tarnishing the cavities which may be silver plated or polished copper, in order to attain high Q.

A terminating load, rated at one-half the highest transmitter power being diplexed in that unit is connected to the reject port. In the event of failure of one or both cavities, half of the power will appear in this load before thermal and other sensors normally turn the affected transmitter off.

The hybrid, sometimes called a *3 dB coupler,* is a four-port device. When power is fed into one port, it appears split 50% between two other ports, thus the name 3 dB coupler. Another very useful feature is that the two split powers are 90° apart in their phase relationship. The fourth port is isolated from the input port typically by 26 to 40 dB. If two of these hybrids are connected back to back, nearly all the power entering the first input port will appear in one of the output ports, and a second port will be isolated.

In TV use, a single hybrid is quite useful in feeding a turnstile batwing antenna, which requires split input power with a phase quadrature displacement. The old square quarter wave TV hybrid diplexers are similar in operation, but they are limited bandwidth devices and not suitable for wide-band use in FM community antenna diplexing.

FM diplexer hybrids are capable of extremely high power. They are usually made of large coaxial components with quarter wavelength coupling bars. Their large physical size and low Q greatly reduces power loss which rarely exceeds 0.05 dB (1.1%) (see Figure 4.12-15).

Many constant impedance diplexers use only reject cavities. Each diplexer contains a broadband input port and an injected frequency port. Nearly all uncombined diplexer power and other undesired products are absorbed by a dummy load connected to the fourth port. The third port goes to the antenna or next diplexer's broadband input port.

A band reject diplexer is shown in Figure 4.12-16. The signal from TX-1 splits equally in hybrid H-1. It

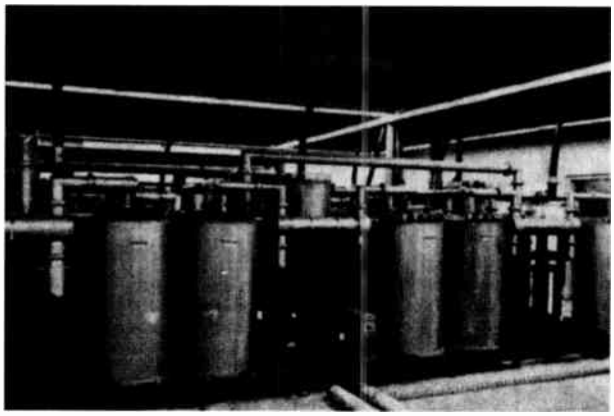


Figure 4.12-15. This Senior Road, Houston, Texas ten-port diplexing system, combines nine FM stations into one community antenna. It is capable of 350 kW total input power. (Courtesy Harris Corp.)

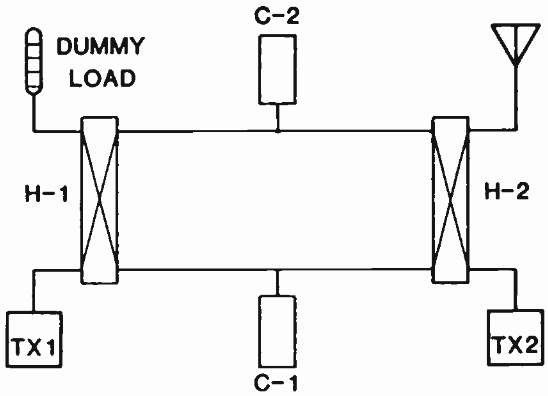


Figure 4.12-16. Basic constant impedance diplexer. See text for explanation of component functions.

passes by the cavities (C-1 and C-2) which are tuned to TX-2 frequency. The two signals of TX-1 combine in the antenna output port of hybrid H-2 since the two inputs to H-2 are equal in amplitude, but 90° out-of-phase, due to the action of hybrid H-1.

The signal from TX-2 splits equally in hybrid H-2 and are 90° out-of-phase with each other. Cavities C-1 and C-2 are tuned to the TX-2 frequency and present a short circuit to it. So the signal is nearly 100% reflected back to the inputs to H-2. These signals are 90° out-of-phase so they combine only in the antenna port of H-2 and go out to the antenna.

TX-1 frequency is not critical and can be any frequency within the FM band, as long as it is removed from the cavity frequencies by at least 1.0 MHz. However, it will work with some makes of reject diplexers when the separation is a minimum of 800 kHz.

The cavities tuned to TX-2 are not critical in their spacing from hybrid H-2. Because of the nature of the two hybrids, the VSWR remains low and is not affected by temperature variations. The isolation also remains high under these conditions.

Another advantage of the constant impedance notch diplexer is that any beat frequency or intermod product generated in the transmitter TX-2 will be absorbed in the dummy load since only TX-2 will be reflected. This is due to the frequency selectivity of C-1 and C-2.

When more than two transmitters are to be combined, the additional diplexers are merely added in series as shown in Figure 4.12-14. Additional diplexers do not affect the performance of the rest of the system, and can be added at a later time if the need arises. Because of their simplicity, constant impedance diplexers are virtually maintenance free.

Branched Starpoint Diplexers

This type does not use hybrids. Each transmitter feeds into a set of branching filters. In each set there is a bandpass cavity tuned to pass the operating frequency and at least one reject cavity tuned to the frequency of the adjacent channel to sharpen the skirt selectivity. If more than two frequencies are being combined, the middle frequencies have reject filters for both their adjacent frequencies.

All the filter sections connect to one common junction, or starpoint, which is the combiner's output. Coaxial cable lengths between each diplexer and the starpoint must be a defined electrical length for proper filtering action and low VSWR. This type of diplexer does not use an absorption load, as the small amount of rejected power is consumed by the rejected frequency transmitter and the antenna.

There is no broadband input, so adding channels requires that balanced modules be added on the output. If there is a failure of one branch, it will usually not take the others out of operation. Since there is no broadband port, the affected frequency cannot be switched as would be done with a hybrid diplexer system. However, the transmitter from the failed branch may be operated into one half of the antenna, while the output of the combiner, containing the unaffected channels feeds the other half, for an ERP reduction of 3 dB.

The principal advantages of a branched system are in space and price. There are fewer components—no dummy loads, no hybrids, and fewer cavities. They are frequently fan cooled, thus permitting the use of smaller cavities. This is all at the price of flexibility and, in some cases, performance.

Cavity Construction

In order for diplexers to work, a frequency selective electrical short circuit is required to be placed between the two hybrids. This is provided by high Q cavities.

During World War II, low-pass and high-pass filters were developed for VHF communications and radar. The need for microwave receivers with greater selectivity led to the development of magnetically coupled, quarter-wavelength long cavities. Improvements since the 1940s have made cavities with excellent loaded Qs and extremely low insertion loss.

Temperature stability improvement is the result of using invar steel, and the high Q results from silver plating and polishing the inside of the cavities. Cavity size for FM use varies due to the resonance mode selected by the design. Round cavities as small as 20 in. (508 mm) in diameter and 30 in. (760 mm) long have been used. Square types as long as 60 in. (1524 mm) using wave-guide modes have been used. Generally speaking, the larger the cavity, the smaller the RF loss through the cavity.

A very high loaded Q is necessary for a frequency spacing of 800 kHz, which is the closest frequency assignment in any specific community. Practical Q values vary from 1,000 to 12,000.

The power dissipated in heat will expand critical parts of the cavity, detuning it if the heat is not efficiently removed. Air blowers, cooling fins, or simple black paint can be used, depending on the amount of heat to be removed.

The cavities contribute most of the loss found in a diplexer. A diplexer with its two cavities tuned to the first possible channel 800 kHz away could have an efficiency of about 95% (0.25 dB loss). Efficiency goes up as the spacing between the passband and the reject frequency increases so that at 1.6 MHz it could be 96% and at 2.3 MHz, about 97%.

An extremely high Q would be excellent on the operating frequency but would not be useful on the FM sidebands. Using more cavities and stagger tuning them increases costs but does not completely solve the efficiency problem.

FM ANTENNA INSTALLATION ON AM TOWERS

The current trend is to locate FM transmitters in places where the best service may be rendered to the most listeners. This usually permits the maximum possible height to be used. Sometimes however, it may be economical and convenient to install the FM antenna on a tower used for AM broadcasting. If the steel AM tower is not base insulated but is grounded and shunt fed, the FM coaxial line may be connected to the tower, without any further problems.

Transformer Isolation

If the AM tower is insulated at the base, an isolation transformer may be used to couple the FM power across the base insulator without introducing objectionable mismatch into the FM antenna feed line. An isolation transformer is especially desirable for feeding high impedance AM radiators or AM radiators which are part of an AM directional antenna system which might be adversely affected by a quarter-wave isolation system.

These transformers have two tightly coupled RF coils which are resonant at the FM operating frequency. An air gap is provided for the AM power to pass through the two resonant loops. The capacity is quite low, resulting in a very high capacitive reactance placed across the tower base insulator.

Figure 4.12-17 shows the internal basic construction of a typical isolation transformer. The insulation for AM under the top of the box may be high density polyethylene, teflon, or fiberglass. The metal top provides a rain shield as well as protection from dust, mud, or snow.

The use of these isolation transformers permits the AM tower to operate undisturbed by the presence of the FM antenna. It also allows the FM coaxial line to be connected in the usual manner, except for the placement of the isolation transformer. These have internal gas blocks and permit the passage of dry air pressure through the transformer via a plastic tube.

In addition to lower cost, the isolation transformer method has another advantage in directional AM tower use. It does not distort the AM radiator current distribution which may adversely affect the AM radiation pattern.

Quarter-Wavelength Isolation

A less popular and older method is to use the technique of quarter-wavelength transmission lines. Simply stated, the opposite end of a shorted quarter-wavelength line has high impedance. This high impedance is placed across the AM tower base and may be successfully used to provide the necessary isolation. It is more difficult to physically accomplish this as the tower should be at least one quarter-wavelength high, and the FM antenna coax line must be insulated all the way down the tower. In practice the insulated part may be as short as 75° of line, as the line hangers and distributed capacity of the line tend to electrically increase the physically shorter line. For best results, the FM line should be placed within the tower body.

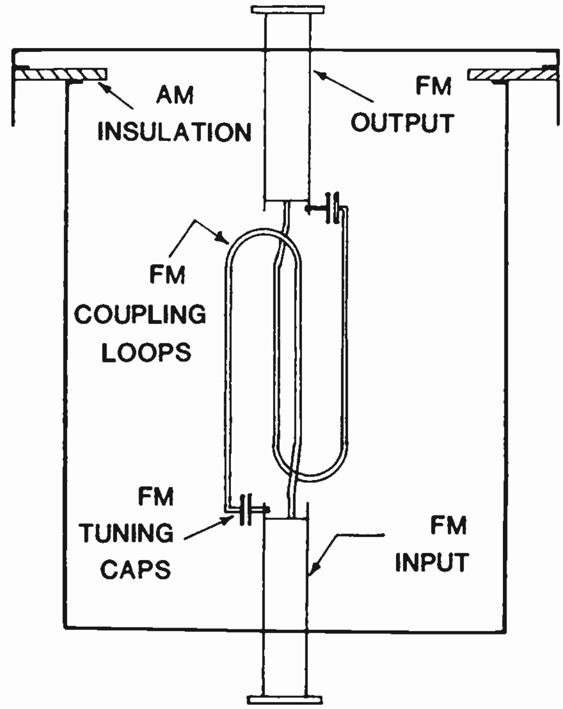


Figure 4.12-17. Typical AM-FM tower isolation transformer, used to decouple FM transmission line on a series fed AM tower.

Guy Cable Considerations

The presence of continuous steel guy cables going through the FM antenna level on a steel supporting tower was studied by Jampro Antenna Company in 1968. It was found that guy cables had an effect of less than 0.6 dB in the azimuth pattern of a H-pol antenna. On V-pol the maximum variation was 1.8 dB on the azimuth pattern. The strongest effect is on CP where the azimuth pattern change was as great as 3.4 dB. The elevation pattern was also affected since the first and second nulls were filled as much as 4 dB.

In addition to pattern anomalies, the steel cables reradiate near the ground. This may cause RF feedback problems in some high power installations with low level audio equipment located in a building near the base of the tower.

Strong currents may be induced when the steel guys are in the immediate vicinity of the radiators. If that guy passes close to the side of the radiator, the field on that particular side of the antenna element will induce currents in the guy wire. The radiator currents can become unbalanced and the impedance of the element is disturbed, changing its VSWR and radiation pattern.

Because of these effects, it is common practice to break-up the guys using insulators, fiberglass rods, or plastic guy cable, within 10 ft (3 m) of the antenna radiators.

When a CP antenna is side-mounted on a guyed tower, the vertically polarized field will have an appreciable component parallel to the guy wire in its aperture and will induce currents in the wire. If the guys are continuous, a progressive wave traveling toward the ground will result, and will radiate most of its energy before reaching the ground. The energy will be radiated in cones concentric with the wire. A small amount of the V-pol power will thus be bled off.

Porcelain Insulators

If the FM antenna is side-mounted on an AM tower, which usually will have metal guys and porcelain break-up insulators, those insulators will probably be spaced several FM wavelengths apart. The induced currents will form standing waves on the sections between insulators and radiate multilobed patterns into space at many angles from the wire axis. If the sections between insulators happen to be of a resonant FM wavelength however, currents in the guy wires and their radiated fields will be considerable.[22] A single isolated piece of guy wire with its ends insulated can only resonate in multiples of one half wavelength, so this spacing of insulators must be avoided. In fact, with the capacitive end loading of the insulators, the resonant length of wire will be somewhat less than one half-wave, so ⅜ wavelengths should also be avoided. A quarter-wavelength is much better, but this would be quite expensive as it requires insulators every 30 in. (762 mm).

Alternatively, the guys through the FM antenna aperture on the tower could be replaced with plastic cable, which is transparent to RF energy.[25]

Figure 4.12-18. A roll of plastic guy cable with eye and jaw end connectors in place. Diameter strength is equal to EHS steel stranded guying cable. (Courtesy Phillystran, Philadelphia Resins Corp.)

Nonconductive Guys

In order to eliminate FM pattern distortion, any guy cable going through the antenna level should be of non-conducting material. Plastic fiberglass (GRP) insulating rods as well as flexible plastic rope covered with a PVC plastic jacket can be used. The black jacket prevents deterioration due to ultra violet sunlight radiation, which may be injurious to the plastic strands of the rope. Plastic rope has been successfully used for more than 25 years.

The idea is to remove metallic RF conducting steel guy cables from within the antenna aperture. The rest of the guy may be of steel construction. The length of the steel guy from its attachment point near the antenna to a point well below the antenna is simply replaced with an equal length of fiberglass rods, or plastic rope.

Plastic rope is available in continuous lengths of up to 1,000 ft (304 m) and kits are available for installing the end fittings. The cable is quite flexible as the Figure 4.12-18 shows a 225 ft (69 m) length coiled up with its end fittings installed at the factory.

The cable may be purchased in strengths exceeding similar diameter steel guy cable. These strengths are shown in Table 4.12-8 for corresponding size of com-

Table 4.12-8
Phillystran type HPTG plastic guys

Outside Diameter		Break Strength		Jacketed Weight		EHS
Inches	mm	Pounds	Kgs	1,000 Ft	300 m	Equivalent
0.20	5.1	4,000	1,815	18	8.2	3/16
0.29	7.4	6,700	3,039	31	14.1	1/4
0.42	10.7	11,200	5,080	55	24.9	5/16
0.46	11.7	15,400	6,975	69	31.3	3/8
0.53	13.5	20,800	9,435	93	42.2	7/16
0.58	14.7	27,000	12,247	115	52.2	1/2
0.63	16.0	35,000	15,876	142	64.4	9/16
0.68	17.3	42,400	19,235	167	75.8	5/8
0.73	18.5	58,300	26,445	195	88.5	3/4

monly used extra high strength (EHS) steel guy wire. Sizes smaller and larger than shown in the table are available.

INSTALLATION PROCEDURES

If the installation is not properly planned and carried out, there may be unwarranted delay and cost associated with putting the FM antenna on its support tower. The following suggestions are offered to avoid unnecessary delays and expenditures.

Planning The Installation

Because of the extremely high cost of rigging services, it is essential to carefully plan the installation. Make sure that all parts are on hand.

The installation of the antenna should be planned by a technically qualified person who must supply accurate tower construction information to the antenna manufacturer. If this information contains errors, these will be carried through the design and fabrication of mounting hardware, and finally show up in the field to plague the installing crew, wasting time and money at every stage of the process.

The station should consider hiring a tower rigging firm that is financially qualified and mechanically well equipped to do the work. A written contract should exist between the station and rigging firm, with a fixed price. The rigging contractor should be licensed as a contractor in your state, and should post a completion bond. He should also supply an insurance policy holding the station harmless, and making the station and its personnel co-insured. Only in this manner will the broadcaster be protected in the event of injury or property damage.

The riggers should be knowledgeable about antennas and coax line, and should inspect the tower and check out the mounting design of the brackets before the full rigging crew arrives.

If any factors are discovered which appear to negate the installation design, contact the factory immediately. Particular attention should be paid to the following:

- Fit of mounting brackets to tower member
- Freedom from interference of the mounts with gussets, leg flanges, guys, and their attachment points, tower face members, and obstruction lights
- Compatibility of transmission line and antenna input coax terminals
- Location of transmission line run relative to antenna input terminals
- Use of fiberglass guys on the tower in the region occupied by the FM antenna (refer to the AM Guying section in this chapter)
- Availability of proper voltage, current, and cable size for deicers if required
- Adequacy of tower to carry the windload placed upon it by the antenna, particularly where radomes are used. This radome/antenna load should be checked by a competent structural engineer, as all antenna installations should be checked. This is usually required by the company carrying insurance on the tower.

Receiving And Unpacking

The boxes are usually numbered and the total number is indicated on each box; contact the shipper if not all boxes are delivered, or if equipment is received damaged. Do not store the material outdoors, boxed or otherwise.

As soon as the antenna is received, open and examine for shipping damages so that any necessary claims may be filed with the shipping company immediately. Check the material against the parts list and installation drawing.

The box with the installation drawing and instructions is usually marked. Open it first, so that the balance of the items may be easily identified and counted. Contact the factory immediately if any material appears to be missing or is damaged during transportation.

Do not call the riggers until all antenna and coaxial line at the site. Otherwise, unnecessary delays and costs may result.

Installation Tips

Broadcast antenna manufacturers furnish detailed installation procedures with their products. These instructions should be closely followed by the rigger. Together they will ensure a perfect installation saving time and money.

The following items are specifically called to the attention of the broadcast engineer (in addition to all those stated before) to permit proper installation and good performance for many years.

- Follow manufacturer's instructions. See that the riggers also read these instructions
- Do not leave antenna parts where rain or moisture can enter. Store indoors and keep units capped as received
- Do not allow dirt or other foreign matter to enter any coaxial part
- Protect all antenna parts from physical damage and abuse
- Hoist antenna members carefully, with a tag line to prevent damage by striking against the tower
- Install on the tower as indicated by the manufacturer's instructions, remembering that bay number 1 is the uppermost top unit
- Riggers should lubricate O-rings with a small amount of silicone grease before mating flanges
- The full complement of flange bolts must be used and they should be as tight as possible
- Tuners or individual element devices, if used, should be adjusted only after the entire antenna and tower installation has been completed
- Rigid transmission lines should be properly installed with two hangers per 20 ft (6 m) length, and with the inner conductor retaining pin on the top of each section

- If semi-flexible cable such as Heliax or Wellflex is used, it should be firmly tied down at least every 5 ft (1.5 m) for 3 in. (76 mm) line, and every 3 ft (1 m) for 1⅝-in. (43 mm) coax line. The line manufacturer's hangers should be used. The line should not be attached to the tower using wrap-lock straps
- After physical installation has been completed in accordance with the manufacturer's recommendations, the main transmission line should be pressurized with dry air through a dehydrator, air pump, or by using dry nitrogen gas. (See the Air Pressurization section in this chapter for more information.)
- Dry air or gas pressure should be maintained at all times. Most antenna warranties are not valid unless this is done. It is the riggers responsibility to make certain that the entire coax and antenna holds air pressure
- The antenna system should be checked by a qualified rigger every time the obstruction lights are replaced, or if lights are not used, at least once a year. The rigger should look for vibration and storm damage, loose or broken coax hangers, and signs of arcing across exposed insulators. A dry rag soaked in 91% isopropyl or other solvent alcohol or equal should be used to wipe clean all exposed insulators in each antenna element. (Do not use carbon tetrachloride!)

STRUCTURAL CONSIDERATIONS

Most FM antennas in the western hemisphere are installed on the sides of a steel tower, between 18 and 60 in. (45 to 152 cm) wide. The antenna and its transmission line together with all mounting brackets introduce wind loading, in addition to their dead weight. The live wind loading is a result of the amount of physical surface presented to the wind. It is sometimes called the *wind catch area*. This consists of either flat or round antenna members, coaxial lines, mounting brackets, and hardware, all represented as surfaces which are exposed to the wind.

The dead weight of the antenna system is fixed and is always present on the tower. The live load is a variable, depending on the wind velocity, and is added to the dead load for the total amount present.

The standard wind load starts with an assumed wind velocity of 87 mph (139 m), which will produce a push of 35 lbs per square foot. (170.8 kgs/sq.m) With lesser wind speeds the wind push is less, and more with higher velocities. Various building codes determine the rated winds to be considered in the design of the tower system. While most of the United States has a 35 lb per square foot minimum rating, some parts of the country have higher requirements due to higher wind velocities. Some insurance companies may require even higher safe wind ratings.

ANTENNA POLE MOUNTING

Nonsymmetrical antennas may also be installed on a round pole, made of various diameters of steel pipe. Several antenna manufacturers supply these as a complete system and will optimize the horizontal plane pattern. The advantage of pole mounting on top of a tower or building is that the pattern may be more easily contoured. This provides more signal in the service area since the antenna orientation is not limited by a fixed triangle formed by a guyed tower.

TRANSMISSION LINE SYSTEMS

Two types of coaxial transmission lines may be used to feed FM antennas. One uses rigid coaxial line sections, each 20 ft (6.09 m) long and requires elbows, flanges, spring hangers and other devices.

The other has a semi-flexible coaxial line which is available in either air or foam dielectric, with fixed hangers. Semi-flexible cable is available on a spool whose diameter depends on the line size. EIA end flanges mate to the antenna flanges as well as other RF equipment. A typical semi-flexible coaxial transmission line layout is shown in Figure 4.12-19.

A supply of dry air should be used with all air dielectric transmission line. Its purpose is to keep out moisture, which may find its way into the line. Dry air keeps the moisture from covering the internal plastic insulation, and thus the potential for arcing is greatly reduced. It also keeps the internal copper from corroding. Copper which has been exposed to moisture will turn dark brown and oxidize after time. Rigid as well as semi-flexible line in the 3 and 3⅛ EIA size which had not been pressurized have been measured with increased attenuation of nearly 4 dB per 100 ft (30.4 m)

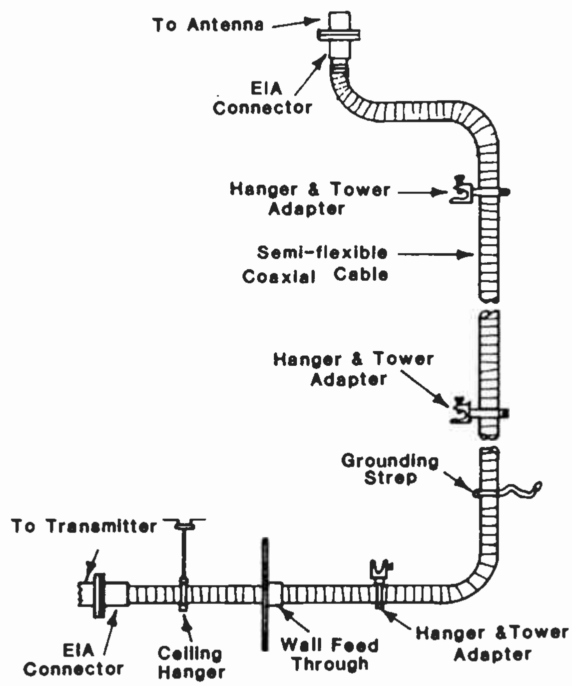


Figure 4.12-19. Typical semi-flexible coaxial cable installation on a tower. Number of various pieces of hardware depend on length and the diameter of the cable.

Table 4.12-9
Coaxial transmission line (Characteristics At 1000 MHz)

Line Type & Number	Nominal EIA Size	Attenuation in dB	100 Ft Eff. %	Safe FM Power-kW
-Foam-				
LDF5-50	7/8	.385	91.5	5.2
LDF7-50	1-5/8	.231	94.8	13.1
-Air-				
HJ5-50	7/8	.373	91.7	6.4
HJ7-50A	1-5/8	.205	95.4	14.2
HJ8-50B	3	.142	96.7	36.8
HJ11-50	4	.115	97.4	56.1
HJ9-50	5	.078	98.2	73.9
-Rigid-				
1213-1	1-5/8	.191	95.7	10.0
1313-1	3-1/8	.096	97.8	40.0
1413-1	4-3/8	.070	98.4	81.0
1613-1	6-1/8	.050	98.8	161.0

at 100 MHz. This is a power loss of 60%. (See the section on Air Pressurization on how this may be accomplished.)

Popular types and sizes of coaxial cable transmission line are shown in Table 4.12-9. The most popular are the semi-flexible air cables.

The safe power ratings shown in Table 4.12-9 are for a perfect VSWR of 1.0:1. It is considered good engineering practice to derate this by dividing the power in kilowatts by the expected VSWR. For example three inch Heliax rated at 36.8 kW at 100 MHz and divided by 1.1 derates it to 33.45 kW. Other increases and decreases of coaxial cable ratings may be found in the catalogs of most manufacturers of transmission lines.

Ratings are for a VSWR of 1.0:1; ambient temperature of 75°F (24°C) with one atmosphere of dry air pressure. Rigid line is rated at 122°F (50°C). The LDF and HJ series are manufactured by Andrew Corp. of Orland Park, Illinois while the rigid line is by Shively Labs., of Howell Labs. Inc. Bridgton, Maine.

AIR PRESSURIZATION

If the antenna is operated without positive pressure of dry air or nitrogen, the manufacturer will not assume responsibility for failure under power. Moisture or even the accumulation of water within the coaxial transmission line is a very serious matter. Its presence causes the VSWR to rise. When a sufficient amount of moisture is present, arcing will take place burning the line or antenna radiating elements. High humidity or moisture will cause the inside of the coaxial transmission line to corrode over time, thereby increasing the line loss. For this and other reasons, the entire antenna system must be dry air pressurized.

After the antenna is installed and the transmission line connected, the system is purged with dry gas or dry air to remove trapped moisture before RF power is applied. A manually opened or pressure actuated purge valve is installed in nearly all FM antennas made by American firms. When the gas pressure is raised to 10 psig (0.68 atmospheres), the automatic pressure relief valve (if the antenna has one) will open up letting moist air out. The complete system purge requires a considerable volume of dry gas.

Before expending this amount, it is good practice to perform a quick check for major leaks. The system pressure is raised to a point below the relief valve setting, such as 8 psig (0.48 atmospheres) the source of supply shut off. A pressure gauge should be installed on the antenna side of the shut-off valve. The pressure, when corrected for temperature, should not fall to less than half its initial value in a 24-hour period.

If the pressure loss is more than this, the system should be checked with a leak detector, or soap suds, to locate the leak. A pinched or missing O-ring is the most common cause for large leaks.

Once the system is known to hold pressure, it should be purged with dry air or gas. Either must be dry enough to have a dew-point well below the coldest temperature expected to be encountered. When using nitrogen, it should be of the oil-dried type, to remove nearly all the moisture from the gas.

Five to eight psig (0.34 to 0.48 atmospheres) should be maintained in the system at all times to ensure that no moisture will be able to enter. Very small leaks will pull in moisture, if the pressure is lower than suggested above, when the transmitter is turned off nightly. This is due to the pumping action due to expanded dry air/gas pressure cooling down, and contracting below the outside air pressure, during cold ambient temperatures.

PROTECTION FROM ICING

High Q antennas are subject to increased VSWR ratios as well as pattern distortion, with light to moderate coatings of ice. Low Q antennas such as the panel type are usually not affected in this manner. Where climatic conditions cause sufficient ice or in some cases snow to affect the antenna's performance, there are two remedies. The radiating element may be covered with a plastic cover, or, it may be electrically heated to melt or prevent the formation of ice on its sensitive surfaces.

Electrical Heaters

By far the more popular method of deicing is to order electrical heaters at the time the antenna is ordered. Electrical deicing equipment is supplied as an option and is factory installed. Kits are furnished for interbay connections, but the broadcaster must supply power from the building to the center of large arrays, or the bottom element on smaller antennas. Local electrical codes of course must be followed.

While a thermostat may be used with small total deicer wattages, a power relay operated by such a device is required and furnished with most electrical deicer kits by the antenna supplier. Due to high power costs, a sophisticated deicer control, which operates when both temperature and humidity conditions produce sleet or icing, is often required.

Most deicers use a resistance heating element which is inserted inside the antenna radiator arms. One manufacturer, however, uses a different method, dropping the 230/240 volts to a few volts with a transformer located at each bay level. The low voltage is passed through the ice sensitive arms of the radiator and connected to the far ends by a heavy teflon coated wire. The current return is by the stainless steel antenna element, whose ohmic resistance is sufficient to produce enough power heat loss to melt or keep the ice off. This method is becoming obsolete as the transformers are expensive and heating costs are rising, as hourly electrical rates go up. The voltage dropping transformers are not as efficient as direct heaters.

A word of caution when selecting a FM antenna with electrical heaters. Some deicers use 1 kW of power for each bay as previously described and increase the wind loading by 225%. Others have a switchable power option feature using 125/500 W per element, with only a 15% increase in windloading, when compared to an antenna without electrical deicers. The continuing cost of electrical deicers is a consideration of the operational cost of the station and should not be overlooked.

Automatic deicers are those with a thermostat for mounting near the antenna for accurate temperature sensing of the actual ambient temperature. The temperature zone of +20°F to +35° (−7°C to +2°C) is generally the most likely icing range, depending on humidity conditions. Deicers should be turned on at +35°F, prior to ice formation, because it is better to prevent icing than to remove it once it is formed. Power should be turned off when the temperature goes below +20°F since ice does not usually form below this temperature. Figure 4.12-20 shows electrical deicers being wired.

Figure 4.12-20. Rigger finishing electrical tie in box for deicers. Looped stainless steel heaters were factory installed in the four arms, plus one in the feed support boom.

Radomes

A radome is a protective dielectric housing for an antenna radiating element. Its function is to protect the antenna not only from ice, but snow and physical damage due to ice dropping from above. Radomes also protect the radiating element from environmental corrosive atmospheres.

The primary purpose of using radomes on FM antennas is to prevent the VSWR from rising with the formation of ice, if the site and height causes icing to occur during the winter months. Ice formation detunes high Q radiators, increases the VSWR, and causes vertical plane pattern changes. Figure 4.12-21 shows a typical radome enclosing a radiator.

Ice may form on the radome but does not particularly affect the operation of the radiator if that ice is kept at least 0.05 wavelengths from the sensitive portions of the antenna element.

Radomes are particularly desirable in heavy icing environments where deicers are not adequate even with very high heat density. They are also useful in protecting antenna elements from falling ice when they are so exposed.

Radomes are cost effective with single channel high Q antennas where electrical deicer heating power costs are expensive. The deicer power cost is a continuing one, while radomes are a one time capital investment, which may be depreciated over time.

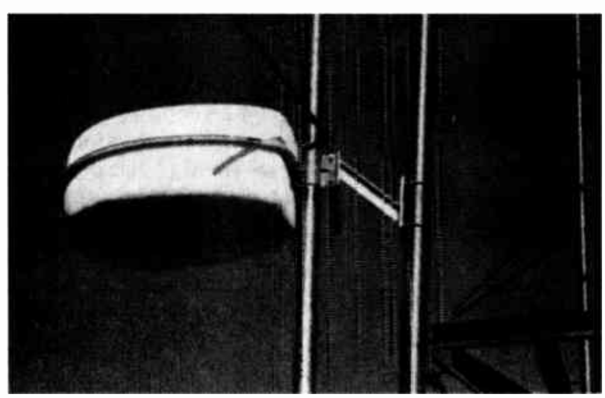


Figure 4.12-21. Radome over radiating element. Transparent to RF, it keeps ice from detuning high Q radiator.

Radomes are generally composed of low-loss dielectrics with low values of dielectric constants and loss tangents. Laminated fiberglass, using glass cloth reinforcements, has a constant of about 4.1 and a loss tangent of about 0.15. Water absorption by the radome increases its dielectric constant and loss tangent. Materials which do not easily absorb water or those treated with a protective gel coat are often used to shed water and prevent the adhesion of ice.

Good radome designs take into consideration operating temperature, a relative humidity of 100%, safe wind pressures, ice, hail, snow loads, rain adhesion, wind and supporting tower vibration, fire retardant plastic, and the ability to safely withstand air contaminants over the useful life of the antenna. All these factors increase the cost, but are necessary for long, useful life. Radome shapes are dictated by the form of the radiating element in most instances.

In all cases radomes are supplied by the antenna manufacturer, and usually supplied in two pieces which are bolted together with stainless steel fasteners.

LIGHTNING

Because FM towers are usually located on high ground, hilltops, or high buildings, they require lightning protection since they are likely recipients of lightning strikes. The type of damage that can be caused by lightning to a FM tower varies. Smaller coaxial lines will usually melt; larger coax (1⅝, 3⅛) will also melt in some cases, and others will conduct the heavy current into the transmitter building to do damage there.

The FM antenna itself may heat, arc, melt, and otherwise be damaged. Holes in the outer conductor, burns, and melting at flanges are common. Teflon or polyethylene insulation will burn, depositing a film of carbon, causing further damage if RF from the transmitter continues after the strike.

Protection of the FM antenna system may be provided to some degree, by taking several precautions. The top of the tower should have a lightning rod, about 1 ft (0.3 m) higher than the uppermost obstruction light part. The FM antenna itself should be firmly grounded to the tower. If the coaxial cable is buried between the tower and the transmitter building, it must be at least six ft away from any tower base grounding system copper wire or strip.[26]

A ground system should be located immediately around the base of the tower. This should have a direct current loss of less than 10 Ω to earth ground. This low resistance may be obtained by using ground wires buried in the soil. Six radials spaced 60° if possible, buried as deep in the soil as possible and running out up to 150 ft (46 m) each, should provide a suitable ground of less than 10 Ω, even if the soil is shallow or rocky.[26]

Guyed tower anchors should also be grounded. This is covered in the chapter on *Design, Erection, and Maintenance of Antenna Structures*. It is important to install the proper number of ground rods and/or copper wire radials in order to obtain a connection to earth ground of less than 10 Ω. In any event these ground rods or radial wires must be tied together with number AWG 4 or larger copper wire, or copper strap 2 in. (5 cm) minimum width. This is to provide for thousands of amperes of current flow for less than one second in the event of a direct lightning hit.

If the FM antenna is located on an AM insulated base tower, then the spark gap should be set at the lowest point providing protection for the highest AM modulation peak voltage.

Another way to protect the FM transmission line isolator (if one is used), as well as the tower and FM antenna, is to use a RF choke across the insulated tower base. This tends to reduce the static build-up voltages due to passing thunderstorm clouds, snow, hail, or dust storms. Arcovers due to these sources usually do not cause damage, but may trip the FM transmitter reflectometer since they will create a current flow through the reflectometer circuitry.

If the base insulated tower supporting the FM antenna is located in an area of regular thunderstorms, another way to protect both antennas from lightning is to ground the AM tower at the base, and shunt feed it. Several excellent methods exist. The folded unipole method not only grounds the tower for lightning purposes, but improves the VSWR bandwidth, so necessary for AM stereo. (See Chapter 4.8, *AM Antenna Systems*, for several suggested methods.)

FM SCA MULTIPLEXING

With a 92 kHz subcarrier and 110% modulation, intermodulation products may be created due to mixing of the various subcarriers with their own harmonics within nonlinear devices. One such nonlinear device can be the antenna system.

Antenna linearity is determined by its VSWR response curve versus frequency. Phase delay in the antenna system is also important. In the past, with 67 kHz being the highest SCA frequency, the ±100 kHz bandwidth was considered sufficient. Now with a 92 kHz subcarrier and 110% deviation (82.5 kHz), the minimum bandwidth is ±130 kHz under 1.1:1 VSWR.[27] (See VSWR Bandwidth in this chapter for more specific requirements and recommendations.)

Tests have shown that 92 kHz is the frequency of choice for a new aural SCA service after 67 kHz.[28] The 92 kHz subcarrier produces lower intermodulation product levels and less interference to the main channel stereo service than 67 kHz.[29] It may be successfully operated in addition to stereo and existing SCA services.

Other nonlinearities in the exciter, and transmitter, plus multipath reception, receiver misalignment, and user mistuning are contributions to the received intermodulation distortion of the baseband signals. In addition these products can cause small levels of audible swishing beat notes in some types of FM receivers.

SPURIOUS FREQUENCIES

Interference to other stations within the FM broadcast band as well as to other services outside the band can be caused by RF intermodulation product energy developed between two or more FM broadcast transmitters. It may be due to coupling through a diplexer or coupling between two antennas. This phenomenon has been well documented.[12]

Detailed information on the susceptibility of various types of transmitters to interference from other collocated transmitters has not been thoroughly investigated. A method has been devised by which the mixing loss between two transmitters can be accurately characterized.

When RF energy from two or more transmitters is combined, new spectral components are produced by mixing the fundamental and the second harmonic of each of them. The dominant intermod product generated by each transmitter is at twice the transmitter's frequency. For example, 101.1 and 102.7 transmitters would produce two intermod signals appearing on 99.5 and 104.3 MHz.

Second harmonic traps or low-pass harmonic filters in the transmission line of either transmitter prior to the diplexer have little effect on the generation of intermod products. This is because the harmonic content of the interfering signal entering the transmitter output circuit has much less effect on intermod generation than the harmonic content within the nonlinear device itself. The resulting intermod falls within the passband of the low-pass filters and outside the reject band of the second harmonic traps, so these devices offer no attenuation to intermod products.

Even the perfect diplexer by its very nature will reflect some of the undesired energy back to each transmitter, generating intermod products. The key to this problem is to keep that undesired power level as low as possible using proper transmitter output circuitry and tight diplexer specifications. Diplexed transmitter installations should be routinely checked for the presence of excessive intermod products.

HARMONIC FILTERS

FCC Rules Section 73.317(d) calls for the harmonics of FM transmitters to be up to 80 dB or more below the transmitter output. This requirement is usually met by using a low-pass filter which passes the station carrier frequency power but attenuates its harmonics.

The transmitter provides some harmonic attenuation of course, and is usually 25 to 38 dB for single ended amplifiers. The worst case harmonic is the third. Harmonic filters by several firms provide a minimum of 50 dB for harmonics from the second through the tenth. Adding the transmitter attenuation to that of the filter normally provides more than the required level.

The high level of rejection is made possible by using high-impedance (inductance) and low-impedance (capacity) coaxial sections for m-derived three to five section filters, with half-pi end sections. Harmonic filters are commonly made in three production schedules. One firm actually adjusts them to the customers operating frequency, so that there are no attenuation gaps in the higher harmonics. They are not tunable outside the factory as the insertion loss, and attenuation along with pass band VSWR are closely related. It requires sophisticated knowledge and equipment to properly adjust.

The insertion loss in the pass band varies from 0.05 to 0.08 dB while the rejection from the second through the tenth harmonic is from 50 to 60 dB, depending on the number of internal filter midsections. The VSWR in the pass band varies from 1.05 to 1.1:1. Harmonic rejection is due to the very high VSWR on the harmonic frequencies which may be as high as 15 to 1. This rejected power is passed back to the transmitter amplifier.

Harmonic filters are available in straight rigid coaxial line sections and may sometimes be pressurized. Power capacity varies from 10 kW for 1⅝ EIA line size to 50 kW for the 6⅛.

ACCESSORY ANTENNA SYSTEM EQUIPMENT

Several other devices are associated with the antenna system. The dry air pressurization of coaxial transmission line was discussed under that heading in this chapter.

Reflectometers

The reflectometer is a device for detecting the ratio of power flow from the transmitter to the antenna (forward) and the rejected power back from the antenna (reverse). A short coaxial line section about 12 in. (305 mm) contains diode detectors, coupling loops and terminations to produce DC current, which drives a suitable VSWR meter.

Reflectometers are wide band devices and therefore must be placed AFTER the harmonic filter. Putting them between the transmitter and the filter causes them to read the rejected harmonic power along with the reflected, thus giving an erroneous reading.

Dummy Loads

A very useful test device in an antenna system is the terminating load. At least one is needed when two amplifiers are combined and fed to the antenna, or when a number of diplexers are used in a community antenna arrangement. Dummy loads are available in several power levels up to 50 kW or more and are cooled by air or water.

RF Switches

Often used with a dummy load, coaxial line switches are available to provide electrical or manual switching of transmitter power to diplexers, antennas, standby transmitters, etc. They typically are not pressurized.

REFERENCES

1. H-pol, V-pol, and E-pol are unofficial abbreviations used for clarity and brevity in this chapter.
2. Broadcasting Yearbook 1991. Broadcasting Publications, Inc. Washington, D.C.

3. Onnigian, Peter, "A Study into the Effects of Vertically Polarized Radiation in FM Broadcasting," Technical monogram. Jampro Antennas, Inc., 1965.
4. Freeman, Roger, "Telecommunications Transmission Handbook," John Wiley & Sons, New York, pp. 180–186, 1975.
5. Head, Howard, "Influence of Trees on TV Field Strengths," *Proceedings of the Institute of Radio Engineers,* Vol. 48, pp. 1016–1020, June 1960.
6. Armstrong, A., "Study of Electromagnetic Wave Propagation at 112 MHz." *Proceedings of the IREE Australia,* pp. 105–110, April 1969.
7. Moeller, Adolph, "Effects of Ground Reflections on Antenna Test Range Measurements." *Microwave Journal,* pp. 47–54, March, 1966.
8. Reed, Russel, "Ultra High Frequency Propagation." 2nd edition, Boston Technical Publishers, pp. 223–238, 1964.
9. Skolnik, Merril, "Radar Handbook." McGraw Hill Book Company. New York, New York, pp. 39–8 to 39–9, 1970.
10. Saveski, Peter, "Radio Propagation Handbook." TAB Books Inc. Blue Ridge Summit, Pennsylvania. pp. 148–159, 1980.
11. Gray, Lawrence, "Radio Transmitters." McGraw Hill Book Company. New York, New York, pp. 181–186, 1961.
12. Mendenhall, Geof, "Study of RF Intermodulation Between FM Broadcast Transmitters Sharing Diplexed Antenna Systems." *Technical Monograph.* Broadcast Electronics, Inc. Quincy, Illinois, 1983.
13. Onnigian, Peter, "Stereo Degradation as a Function of Antenna System VSWR" Audio Engineering Society Annual Meeting, Audio Engineering Society, New York, New York, 1976.
14. Mendenhall, Geof, "The Composite Signal—Key to Quality FM Broadcasts." *Technical Monograph.* Broadcast Electronics, Inc. Quincy, Illinois, 1984.
15. Cohen, Jules, "Proposed Solutions, Channel 6—Educational FM Broadcast Interference Problem." *Proceedings of the 38th Annual Broadcast Engineering Conference,* NAB, Washington, DC, 1984.
16. Onnigian, Peter, Circularly Polarized Antenna, U.S. Patent 3,541,470.
17. DuHamel, Ray, "TV and FM Transmitting Antennas." *Antenna Engineering Handbook.* Johnson and Jasik, 2nd edition. McGraw Hill Book Company, New York, New York, Chapter 28, pp. 8–9, 1984.
18. Bock, E., "Sleeve Antennas." *VHF Techniques, Vol. 1,* McGraw Hill Book Company, New York, New York, Chapter 5, pp. 119–137, 1947.
19. Knight, Peter, "Re-radiation from Masts at Radio Frequencies." *Proceedings of the IEEE, Vol. 114,* pp. 30–42. January, 1967.
20. Jampro Antennas, Inc. International Technical Communication. Sacramento, California.
21. FCC Public Notice. "Criteria for Licensing of FM Broadcast Antenna Systems." Notice 84–437 25004, September 14, 1984, Federal Communications Commission, Washington, D.C.
22. Rudge, A.,"Handbook of Antenna Design." Vol. 2, Peter Peregrinus Ltd., London, England, pp. 917–922, 1983.
23. Onnigian, Peter, "Multi-Station FM Antennas." Paper presented at the 23rd Broadcast Symposium, IEEE Group on Broadcasting, Washington, DC 1973.
24. Fisk, Ronald, "Design and Application of a Multiplexed Nine-Station FM Antenna, Senior Road Tower Group." *Technical Monograph,* Harris Corporation, Quincy, Illinois, 1983.
25. Gregorac, L., "Electrical and Mechanical Analysis of Plastic Guys of Broadcast Towers." *Technical Monograph.* Radio-Television Ljubljana, Yugoslavia. 12 pp. 1973.
26. Marshall, J.L., "Lighting Protection." Canadian Broadcasting System, John Wiley & Son, New York, New York, pp. 53–54, 1973.
27. Kean, John, "Distortion of FM Signals Caused by Mismatched and Limited Bandwidth Transmitting Antennas." National Public Radio, Washington, D.C. pp. 37–42, 1984.
28. McMartin, Ray B., "Super Eight" *Proceedings of the 38th Annual Broadcast Engineering Conference,* NAB, Washington, D.C. pp. 160–166, 1984.
29. Denny, Robert, "Report on SCA Operation." *Proceedings of the 37th Annual Broadcast Engineering Conference,* NAB, Washington, D.C. pp. 187–196, 1983.

4.13
FM COMBINING SYSTEMS

ROBERT A. SURETTE
SHIVELY LABS, BRIDGTON, ME

INTRODUCTION

Transmitting several frequencies from a single broadband antenna system requires the use of a combining system, or combiner, composed of RF filters and an interconnecting transmission line. In general, a combiner can be categorized as one of three types: *runout*, *branched* (star point) or *balanced* (constant-impedance). Any of these types may employ band-reject (notch) or band-pass filters. This chapter discusses the use of filters and other components in FM combiners.

Applications

For years, the FM channel was relatively simple, with only the main channel audio (L + R and L − R) and a stereo pilot. In the mid-1980s, the 67 kHz (Subsidiary Communications Authorization (SCA) became more widely used, then the 93 kHz SCA followed. The 1990s brought Digital Audio Broadcast (DAB).

Each of these developments has widened bandwidths. Updated technologies have been needed to provide the required amount of isolation from other frequencies, while passing a wider bandwidth within each FM channel.

Why Combiners Are Used

As populations migrate to the suburbs, it has become more desirable to construct large broadcasting facilities which can reach these heavily populated areas from more central locations. Of course, these prime locations have become more valuable, so it makes sense to use each to its fullest potential. This can best be done by sharing a transmitter site and a common antenna among several users. To accomplish this, the broadcast industry uses combiners of various types and sizes. For example, in San Francisco (Mt. Sutro), Toronto (CN Tower), Montreal (Mt. Royal), New York City (Empire State Building), and Chicago (John Hancock Building), huge towers or towers on skyscrapers have been used to consolidate as many broadcasting facilities as possible, including VHF-TV, UHF-TV, FM and land mobile communications services. This approach has proven very effective, not only in using real estate economically, but spreading the tower costs over many users. In fact, the CN tower carries virtually all the services in the Toronto area. These installations, of course, employ combining systems to prevent interference among services.

Group ownership of FM stations in a market has led to a proliferation of combined stations. And with the implementation of DTV systems, FM stations are being forced off existing TV towers, making it even more imperative that they share tower space, and increasing the demand for combined systems.

FCC Isolation Requirements

Why not just merge two or more signals in a tee and send them to a common antenna? Because unless they are stopped by some means, the signals will backfeed into each other's transmitters, where spurious intermodulation products (spurs) will be generated within the final amplifier stages of the transmitters, depending on the spacing of the stations. These spurs will then be transmitted up to the antenna and broadcast, causing interference to other FM stations, TV Channel 6 and the aviation band—so care must be taken to prevent them. To keep a transmitter from generating spurs, it must be isolated from unwanted signals by at least 40 dB. Therefore, the basic design criterion for a combiner is to provide 40 dB of isolation, plus a 6 to 10 dB safety margin. The closer the spacing of the frequencies between transmitters, the more filtering is required to attain the 46 to 50 dB of isolation. The farther the spacing, the less filtering is required.

FCC Rule 73.317(d) specifies that spurs more than 600 kHz removed from the carrier be attenuated 80 dB or (43 + $\log_{10}$ (power in watts), whichever is the lesser attenuation) below the carrier frequency.

Modern FM band-pass combiners yield a total isolation of 75 to 80 dB, and when added to transmitter losses of 13 dB, ensure that the 80 dB FCC spur rule is not just met but exceeded.

History of FM Combiners

Dr. Andrew Alford designed several of the early combiner systems in the United States. A major installation was a multichannel system on the Empire State Building in the mid-1950s. This system used pairs of quadrature hybrids with a phase delaying cavity in one leg of each hybrid pair. In the 1960s (for the John Hancock Building in Chicago) Dr. Alford designed a runout combiner system, again using quadrature hybrids. However, the phase matching was accomplished with coaxial runout lines instead of cavities, to combine 10 stations into one broadband antenna. Both of these early combiners have since been replaced by balanced bandpass systems.

In the 1970s, RCA built a three-station runout combiner with the antenna on top of the tallest building (then) in downtown Miami. Unfortunately, a few years later, an even taller building was built across the street! This system was then abandoned and the stations moved to a modern high powered, expandable balanced combiner system, using band pass filtering and, for the first time, group delay equalization modules, in the mid 1980s.

The same purpose has been accomplished in other ways. The Sears Building in Chicago carries a multistation antenna system where, instead of a combining system with a common feed, each station feeds a single-level panel antenna. The antennas are mounted closely together on the tower, well within each other's zone of interaction, so to prevent interference from the others, each transmitter feeds through a three-cavity band-pass filter. To help reduce the interactions, antennas whose frequencies are close-spaced were physically separated as far apart as possible.

In other parts of the world, notably Europe, higher population density and government ownership of broadcasting have led to much earlier adoption of combiner technology, although at relatively low power levels. Many large combined sites have existed for years, such as at the Eiffel Tower and various BBC facilities in Britain.

COMPONENTS OF COMBINERS

Tee or Star-Point Junctions

A *tee junction*, shown in Figure 4.13-1, is a coaxial component that provides for two RF signals to flow into a common path; a *star-point junction* is a tee with more than two input paths. This basic coaxial component is one of the building blocks of a branched combiner.

Resistive Loads

Resistive loads, often called *dummy* loads, are used in many applications and can be manufactured in many sizes and impedances depending on the requirements. In a dummy load, incoming power is absorbed and converted to heat. The heat must then be dissipated to the surrounding air, so the power rating of a dummy load is determined by the size of the resistor and the amount of heat that can be dissipated before the resistor overheats and fails. If enough resistors can be chained together with enough cooling, they can dissipate almost an unlimited amount of RF energy. Most coaxial systems use either 50 or 75 Ω dummy loads, depending on the base impedance of the system.

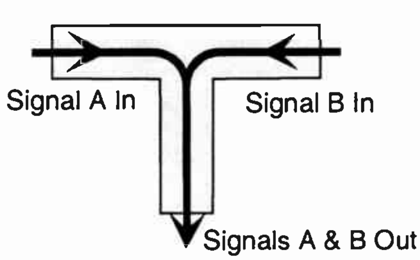


Figure 4.13-1. Tee junction schematic.

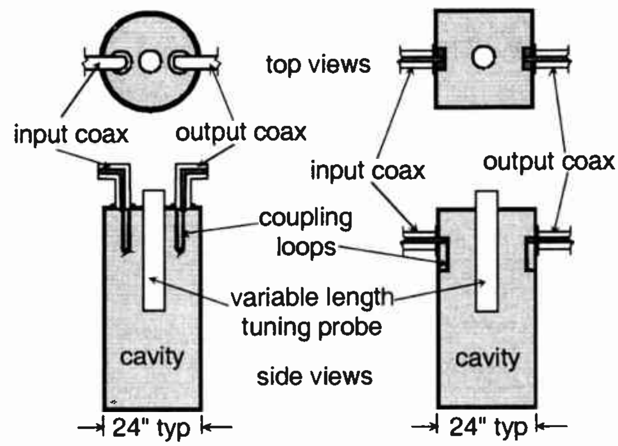


Figure 4.13-2. Band-pass cavity filter designs.

Filters

Filters sort RF frequencies, attenuating some while allowing others to pass readily. Depending on the design, a filter may either *attenuate* (band-reject type) or *pass* (band-pass type) a narrow bandwidth.

The performance of a filter can be expressed in several ways. Common methods include plots of transfer efficiency (frequency response and insertion loss), impedance (Smith Chart) or VSWR, and phase delay (group delay), all as a function of frequency.

Band-Pass Cavity Filters

Figure 4.13-2 shows the basic mechanical configuration of a band-pass cavity filter. When RF energy is applied to the input coupling loop, the loop inductively couples the energy into the cavity. Energy is transferred through the cavity and inductively coupled to the output coupling loop.

The resonant frequency of the cavity is tuned by adjusting the length of the tuning probe. The transfer of energy is maximized at the resonant frequency. Therefore, a filter of one or more identical cavities can be used to attenuate frequencies other than the resonant frequency.

Frequency Response

Energy transfer through a band-pass filter is highest, or least attenuated, at the resonant frequency, and drops off at frequencies above and below that frequency. This *frequency response* is the fundamental property that enables a filter to sort frequencies.

If it were possible to design an ideal filter, its frequency response plot would be as shown in Figure 4.13-3. Response would be flat within the pass band, with a vertical roll-off at the edges of that band.

Curve A of Figure 4.13-4 shows the frequency response of a real-world single-cavity band-pass filter. Note that the energy transfer is highest at the resonant frequency (f_0). The curve is not perfectly flat within the pass band, and drops off gradually away from f_0.

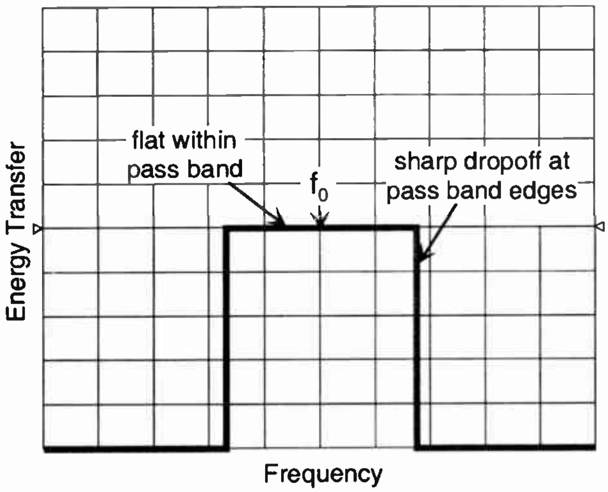


Figure 4.13-3. Frequency response for ideal filter.

Insertion Loss

Even at the resonant frequency, energy transfer is not perfect; some energy is lost along the way. The efficiency of a filter at the resonant frequency is expressed (inversely) as *insertion loss;* that is, the loss of energy at that frequency. For a cavity filter, the lost energy is converted to heat and dissipated in the metal surfaces of the cavity. A cavity that is larger in size is more efficient, in that it will provide a lower insertion loss at the resonant frequency with comparable frequency response. Adjustment of the coupling loops also affects insertion loss (curve B of Figure 4.13-4 shows the effects of coupling loop adjustment).

Our theoretical ideal filter would show no insertion loss in the pass band.

Group Delay

The signal takes a finite amount of time to pass through the cavity, and just as more energy is lost, so is more time taken at non-resonant frequencies. Figure 4.13-5 is a plot of time versus frequency and shows that as the frequency changes further away from f_0, the signal takes more time to pass through the cavity. This is termed *group delay difference,* or group delay. Excessive group delay within the pass band results in signal distortion.

An ideal filter would have no group delay difference; that is, the curve would be a horizontal line, at least across the pass band.

Today's complex broadcast signals use the full channel bandwidth, so it is important to limit group delay. Modern group delay equalization devices compensate for this type of distortion.

A group delay equalizer consists of a *hybrid* and two open-ended band-pass filters, which selectively delay the carrier frequency relative to the sidebands (Figure 4.13-6).

When a group delay equalizer is cascaded with a band-pass filter module, it results in minimal group delay across the pass band, as shown in Figure 4.13-7.

In FM, a group delay equalizer can be installed either at the combiner input, using high-power components, or between the transmitter's exciter and the intermediate power amplifier (IPA), using similar low-power components.

Impedance

Current flow in any RF circuit must overcome resistance, capacitive reactance and inductive reactance. The vector sum of these is termed *impedance.* Because this is a complex function, it may only be fully represented on a complex diagram known as a *Smith Chart.* A full discussion of Smith Charts is beyond the scope of this

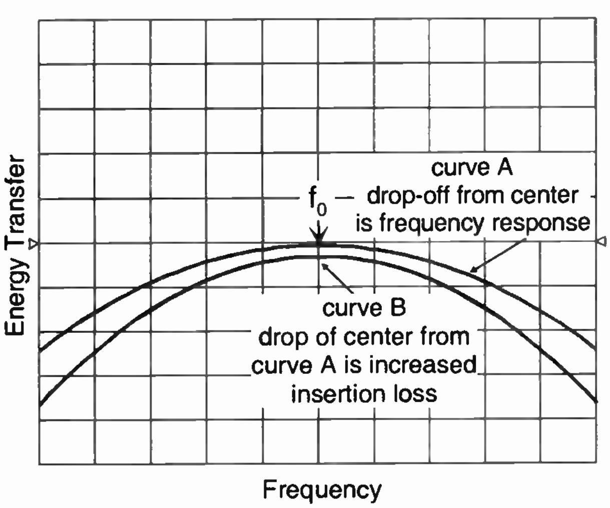


Figure 4.13-4. Frequency response and insertion loss, single-cavity band-pass filter.

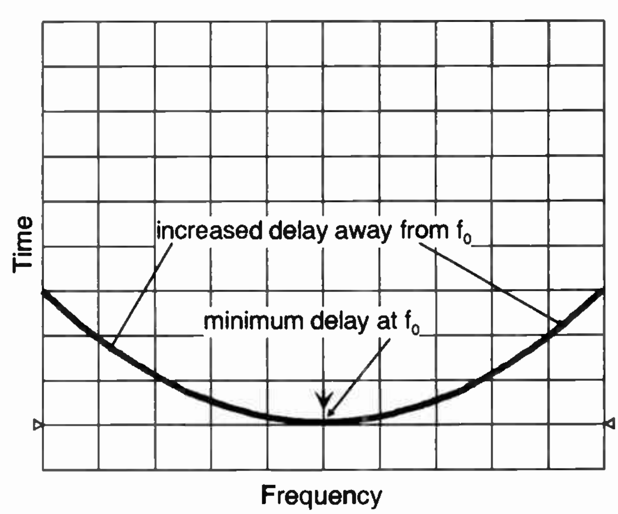


Figure 4.13-5. Group delay for band-pass filter.

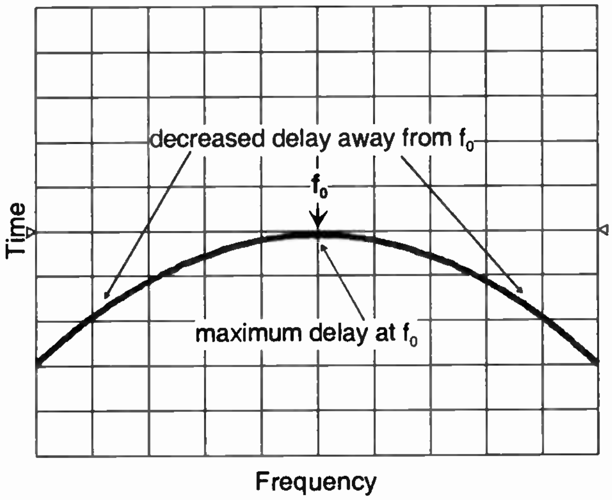


Figure 4.13-6. Group delay curve for a group delay equalizer.

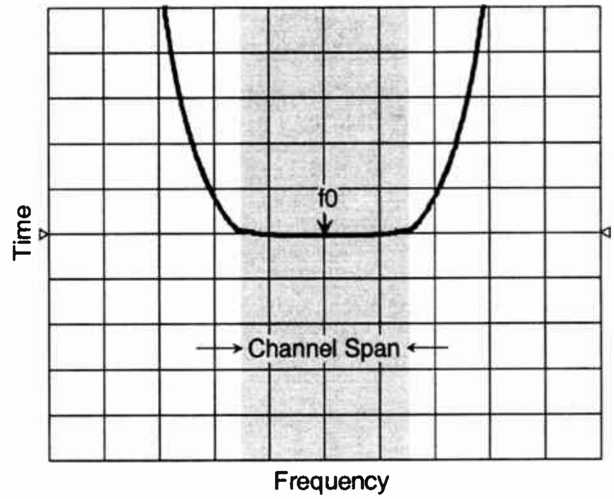


Figure 4.13-7. Group delay curve for a band-pass filter balanced combiner with group delay equalizer.

chapter,[1] but explanation of a few basic features will aid in the understanding of filter performance and tuning.

Figure 4.13-8 shows a Smith Chart. The center horizontal axis (A) represents a state of pure resistance. In a properly tuned system, this state exists at f_0, the resonant frequency, where the inductive and capacitive components cancel each other out. The center point on line A represents the resistive value characteristic of the system being considered (typically 50 or 75 Ω). To the left, the resistive value decreases, approaching a short circuit (0 Ω); to the right, it increases, approaching an open circuit (∞ Ω).

The region above the horizontal axis represents a state when the vector sum of the circuit is inductive in nature. Conversely, below the axis, the circuit is capacitive. Any point on the chart may be expressed as R ± jX, where "R" is the resistive component, "j" is a constant and "X" represents the magnitude of the net inductive or capacitive component of the circuit.

An ideal filter (assuming a 50 Ω system) would be plotted as a dot at the center of the chart, representing a pure 50 Ω resistance throughout the pass band.

Figure 4.13-9 shows the Smith chart for an optimally-tuned single-cavity band-pass filter (50 Ω characteristic image dome). At the resonant frequency f_0, the impedance is pure resistance and 50 Ω at chart center. If the frequency is changed away from f_0, the inductive and capacitive components grow, forming a vertical arc.

The small circles (beads) on the curve indicate the pass band. Usually, a range of ±150 kHz is considered an acceptable pass band for an FM filter system.

[1] *See* Carron, Wilfred N. *Antenna Impedance Matching.* American Radio Relay League (ARRL), Newington, CT, 1989, *or* Besser, Les. *The Smith Chart and Its Applications.* Besser Associates, Los Altos, CA, 1992.

An ideal cavity filter would have a 50 Ω impedance (unity VSWR), no insertion loss, no group delay, and flat frequency response within the pass band. As Figures 4.13-4, 4.13-5 and 4.13-9 show, actual cavities do not meet these ideals. Reaching the aforementioned characteristics is a goal, within the limits of practicality, of filter system design.

If circles were drawn around the center point, each would be a locus of points of equal VSWR; for example, circle B in Figure 4.13-8 represents a VSWR of 1.1: 1. Points within the circle then represent conditions of VSWR less than 1.1:1.

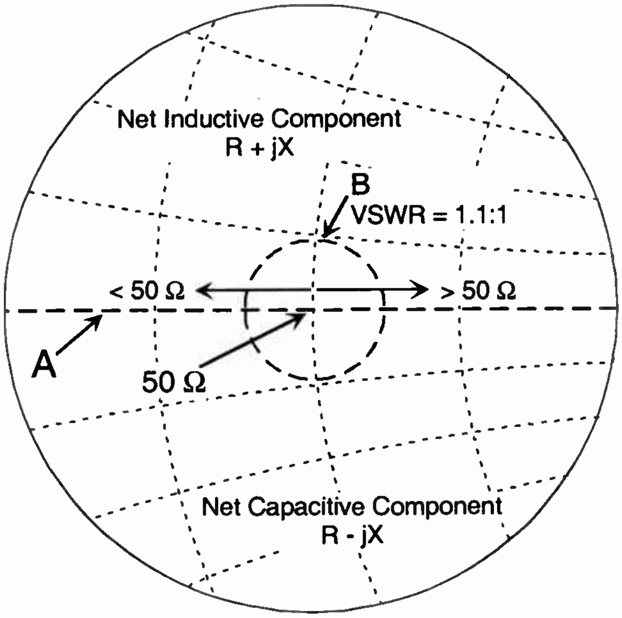


Figure 4.13-8. Smith Chart components.

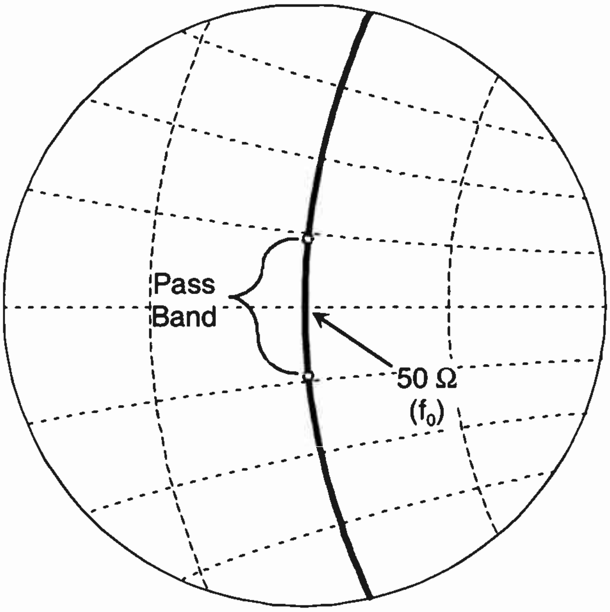


Figure 4.13-9. Smith Chart for a single-cavity band-pass filter.

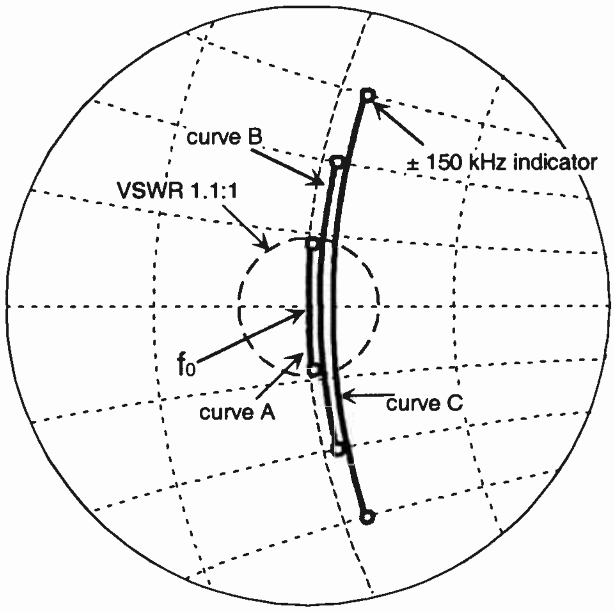


Figure 4.13-10. Impedance diagram, single-cavity band-pass filter, showing effects of loop adjustment.

Single Band-Pass Cavity

The physical size of the cavity is established for the purpose of power handling. The placement, size and rotation of the coupling loops are then chosen to provide coupling which will optimize the cavity performance for a given application.

In Figure 4.13-4 while the center of curve B has dropped below curve A, the outer edges have dropped even more. Although there is increased insertion loss at f_0, there is gained increased *isolation* away from f_0. This is often a favorable trade-off.

Figure 4.13-10 is an impedance diagram showing the same process. Curve A (truncated for emphasis) is an optimally tuned cavity. The curve passes through the center of the chart ($R = 50\ \Omega$), and the entire 150 kHz pass band (between the beads) is within the circle representing VSWR = 1.1:1.

As the coupling is adjusted to achieve increased isolation, (curve B) and still more isolation (curve C), the center of the curve moves into the $R > 50$ area to the right of chart center. In addition, the beads representing ± 150 kHz move outward, well outside the 1.1:1 VSWR area. The trade-off between increased isolation and decreased transfer efficiency can be easily seen.

Generally, a filter system is considered adequate if it provides a VSWR of 1.1:1 over a frequency range of ± 150 kHz. This is termed the bandwidth of the filter system. In most cases, a single band-pass cavity will not yield this much bandwidth. To increase the isolation and improve bandwidth, a second cavity may be cascaded with the first, as shown in Figure 4.13-11.

When two identical cavities are coupled 1/4-wave apart, the impedances superimpose themselves as shown in Figure 4.13-12. Note that Figure 4.13-12 shows two Smith Charts superimposed 180° apart. The beads representing the bandwidth fall on a VSWR circle of about 1.3:1.

When their impedances are added together mathematically, due to phase cancellation, the bandwidth improves to about 1.1:1 VSWR as shown in Figure 4.13-13.

Curve A of Figure 4.13-14 shows the frequency response of the two-cavity filter.

When still more isolation is required, more cavities can be added. Curves B and C of Figure 4.13-14 show the frequency responses of a three-cavity system and a four-cavity system, respectively. As more cavities are added, the curve becomes squarer—flatter across the pass band, with a sharper roll-off; that is, it starts to approach our ideal filter, shown in Figure 4.13-3.

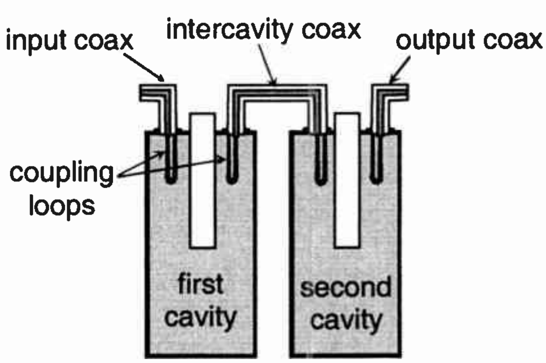


Figure 4.13-11. Loop-coupled cavities.

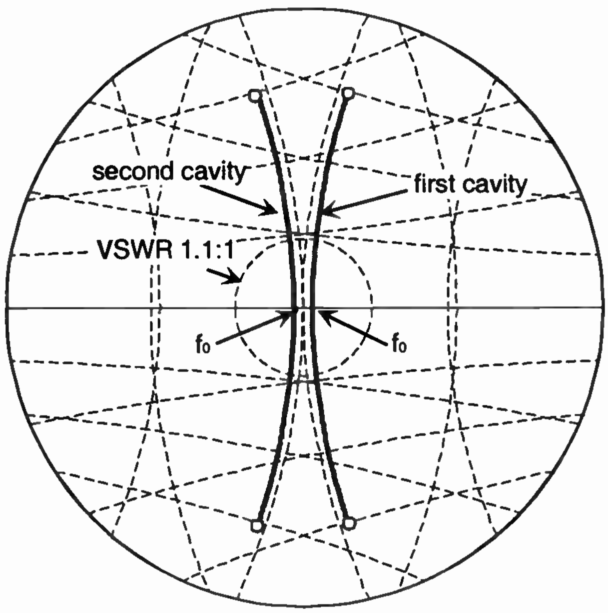


Figure 4.13-12. Impedance diagram, superimposition of two identical cavities, λ/4 apart.

Curve A of Figure 4.3-14 shows the frequency response of the two cavity filter.

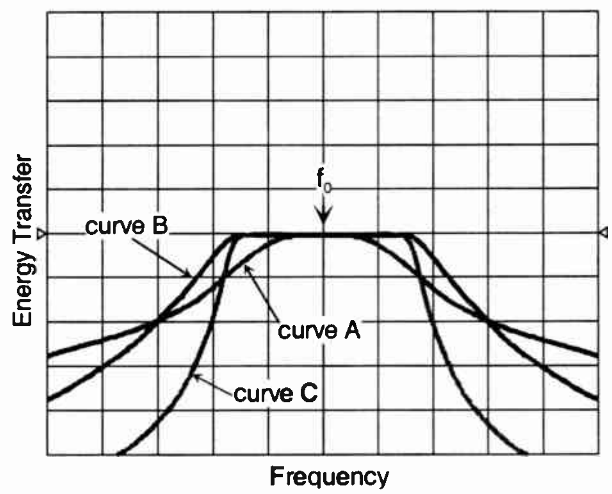


Figure 4.13-14. Frequency response diagrams, (a) two-, (b) three-, and (c) four-cavity band-pass filters.

Figure 4.13-15 shows Smith Charts for a three-cavity system and a four-cavity system. Note that the beads indicating the ±150 kHz points are well within the 1.1:1 VSWR circle.

Tuning Considerations

It is important to remember that filters should always be designed for best real-world *overall* performance. A little performance may have to be sacrificed in one parameter to improve performance in another.

As cavities are added, frequency response is improved—but at the cost of increasing group delay. In a four-cavity system, the group delay curve becomes so steep it is unacceptable (Figure 4.13-16).

Therefore, tuning can be modified to decrease group delay to an acceptable level, as shown in Figure 4.13-17.

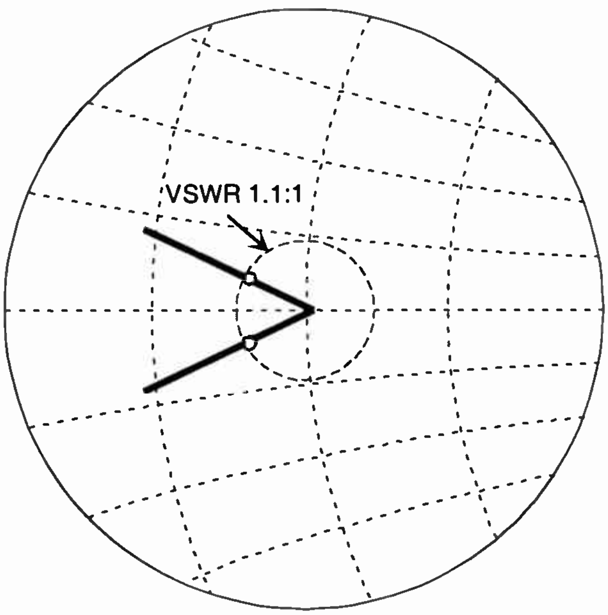


Figure 4.13-13. Impedance diagram, two-cavity band-pass filter.

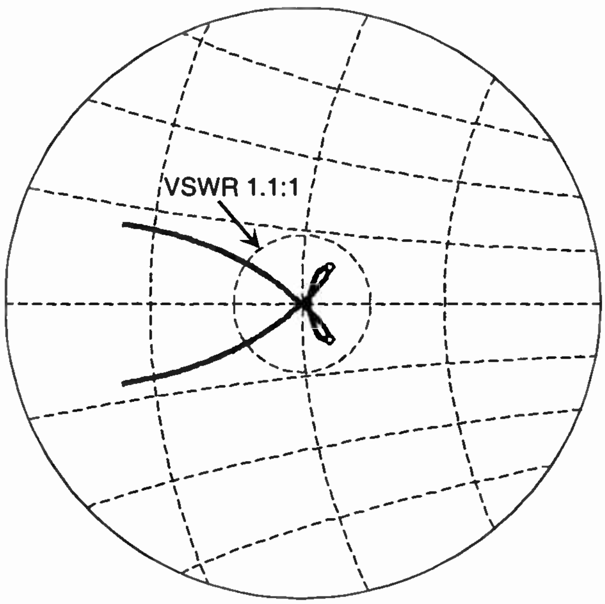


Figure 4.13-15(a). Impedance diagram, three-cavity band-pass filters.

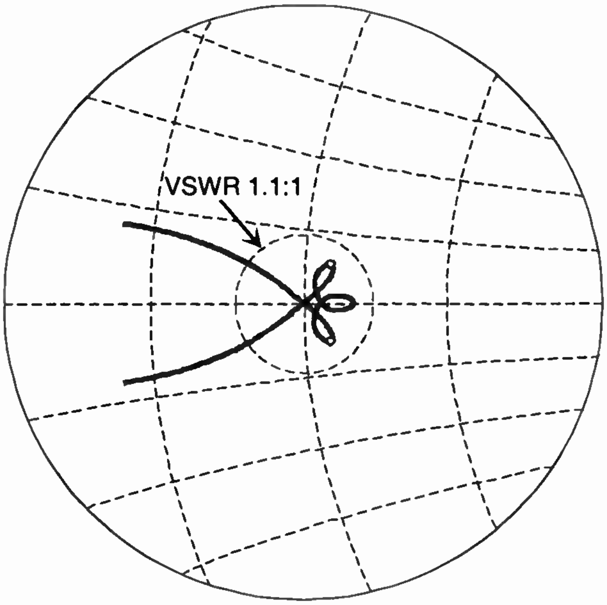


Figure 4.13-15(b). Impedance diagram, four-cavity band-pass filters.

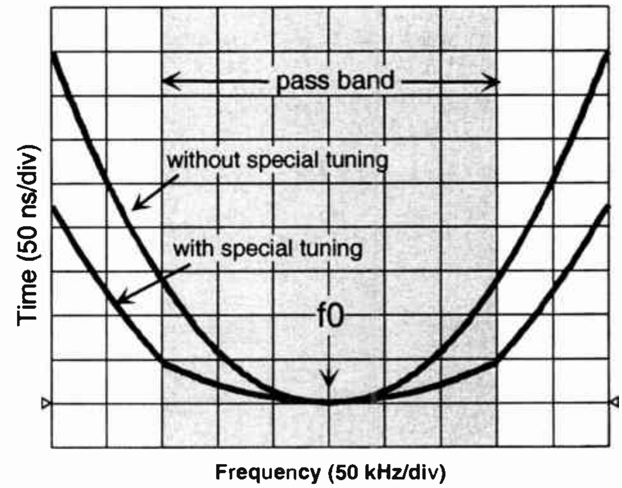


Figure 4.13-17. Group delay comparison, four-cavity band-pass filter, tuned to reduce group delay versus not specially tuned.

This adds some minor distortion to the frequency response (Figure 4.13-18) and considerable distortion to the impedance diagram, as shown in Figure 4.13-19.

Although none of the individual parameters is optimized by itself, the overall performance of the filter is optimized and acceptable. A four-cavity band-pass filter is usually as large a filter system as is needed for most high isolation applications.

Mechanical Constraints

In order to obtain the optimum mathematical cancellation show in Figure 4.13-12 and 4.13-13, the cavities must be spaced at 1/4 wavelength. As the frequency increases, the electrical wavelength decreases—therefore the physical length of the intercavity coax must be shortened. At the higher frequencies of the FM band, the large cavities used for high power applications are difficult to link together, because the cavities themselves approach 1/4 electrical wavelength. As a result, when the intercavity coax is added, the electrical spacing is longer than 1/4 wavelength. In this case, the coupling loops must be manipulated to compensate for the extra length, so that the impedance bandwidth of the cavities is maintained.

Iris-Coupled Cavities

The spacing problem can be prevented by building the cavities contiguous to each other and coupling

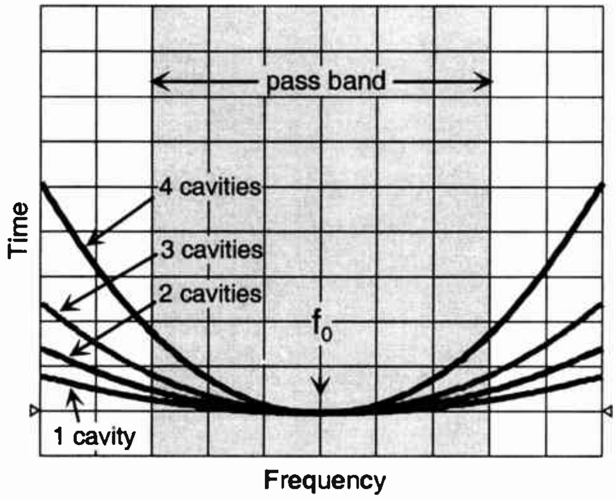


Figure 4.13-16. Group delay comparison, one through four-cavity band-pass filters.

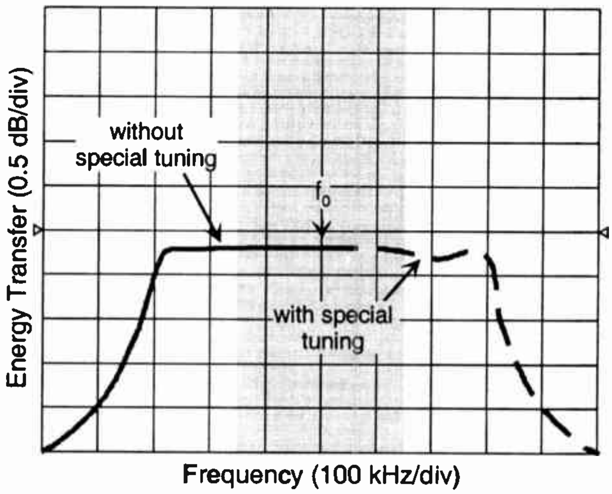


Figure 4.13-18. Frequency response comparison, four-cavity band-pass filter, tuned to reduce group delay versus not specially tuned.

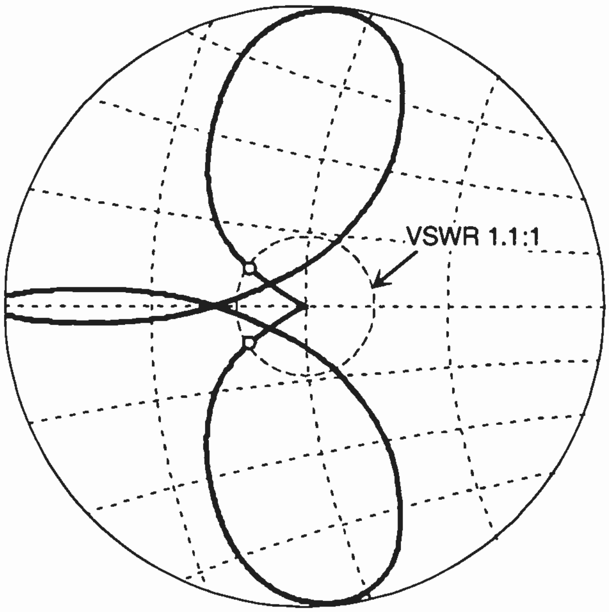


Figure 4.13-19. Impedance diagram, four-cavity band-pass filter tuned to reduce group delay.

them through an iris or slot in the wall between them, as shown in Figures 4.13-20 and 4.13-21. The electrical 1/4 wavelength spacing is maintained by the coupled fields between the cavities.

Band-Reject or Notch Filters

There are several ways to design a notch filter (Figure 4.13-22), but they all accomplish the same purpose. In one form, a notch filter is like a band-pass cavity, but with only an input coupling loop. The cavity is mounted off the transmission line by means of a matched tee. Other designs employ some form of capacitive coupling into the cavity and away from the main transmission line.

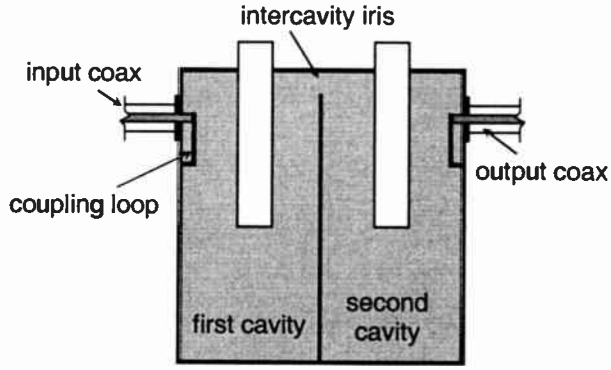


Figure 4.13-20. Iris coupled cavities.

Instead of passing the RF energy *through* the cavity, a notch filter provides a path which *removes* the tuned frequency from the system, allowing other frequencies to pass with minimum loss. Compare the response curve of a single-cavity notch filter, Figure 4.13-23, to that of a single band-pass cavity (see Figure 4.13-4, curve A). The impedance plots of the two filters are basically the same (Figure 4.13-9). When a single notch cavity is used, an impedance matching network is added to the filter to improve the impedance bandwidth.

Multiple Notch Cavities

When more isolation is needed, two notch cavities are coupled in sequence, and if 1/4-wave spacing is used, the impedance bandwidth is basically the same as for a two-cavity band-pass filter (Figure 4.13-11).

When two notch filters are used in sequence, the resonant frequencies of the cavities may be identical, yielding a response curve with a very deep narrowband notch, as shown in Figure 4.13-24, or they may intentionally be staggered, to give a broader notch response, as shown in Figure 4.13-25.

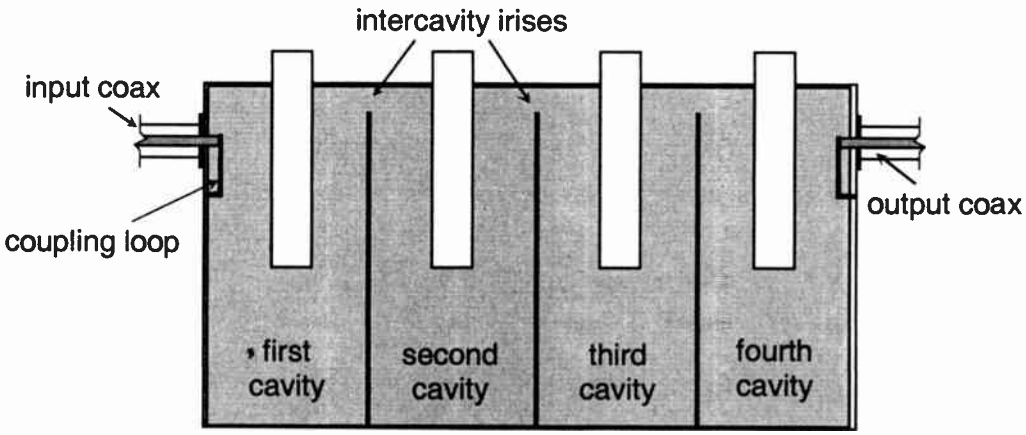


Figure 4.13-21. Four-cavity iris coupled filter.

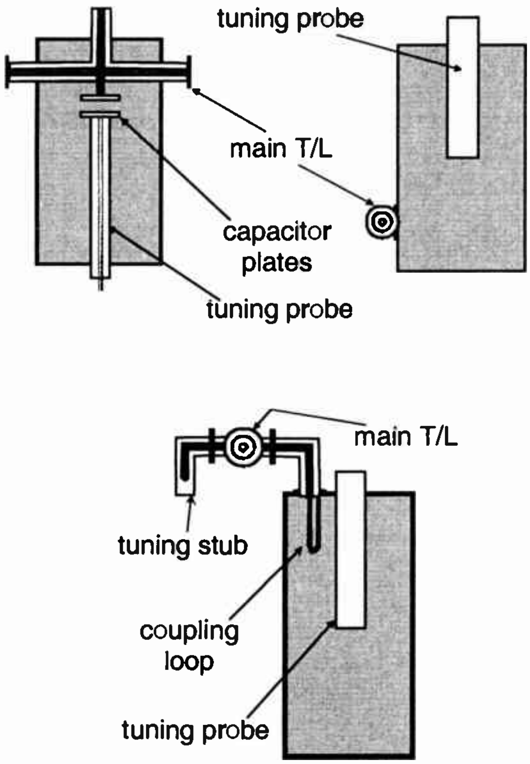


Figure 4.13-22. A variety of band-reject filter styles.

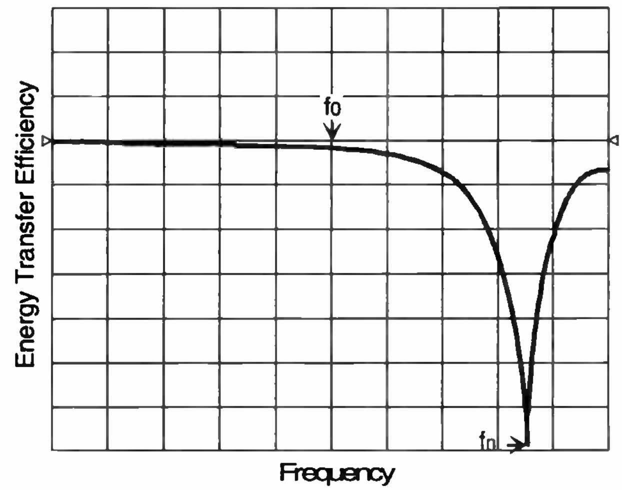


Figure 4.13-24. Frequency response, two-cavity band-reject filter with identical resonant frequencies.

Notch Cavity Group Delay

The group delay plot for a notch cavity is quite different from that of a band-pass cavity (see Figure 4.13-26). However, it is likewise damaging to signal quality.

No practical device has been marketed to equalize the group delay of a notch cavity system. This has not been a major issue, since at about the same time that group delay was recognized as an issue mid-1980s the industry was turning towards band-pass filtering anyway.

Modern Choices

Modern combiner systems are all based on band-pass filters, with limited use of reject cavities to enhance isolation of specific frequencies, for the following reasons:

- Broader response profile. Modern combiner systems must be as broad band as a filter system can be. From the previous illustrations above, it is obvious that the band-pass filter gives a broader band response across the pass band. Furthermore, group

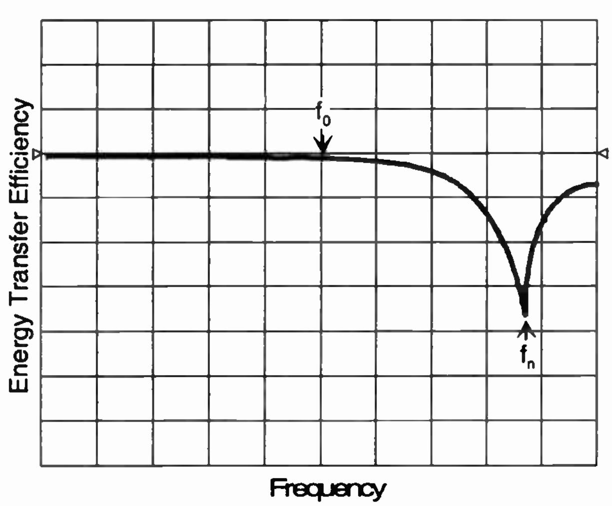


Figure 4.13-23. Frequency response, single-cavity band-reject filter.

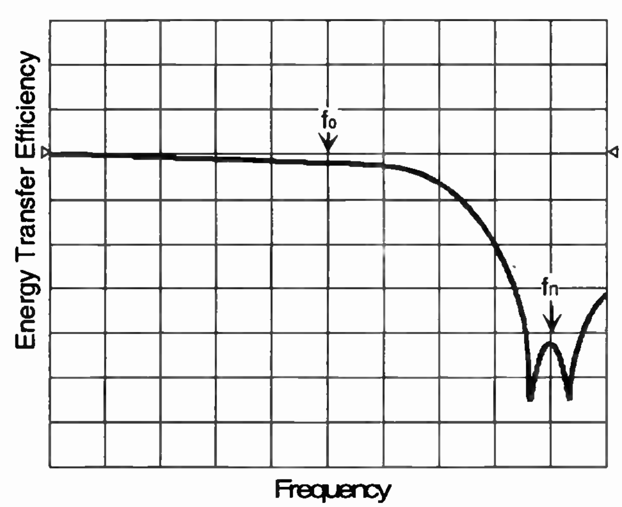


Figure 4.13-25. Frequency response, two-cavity band-reject filter with staggered resonant frequencies.

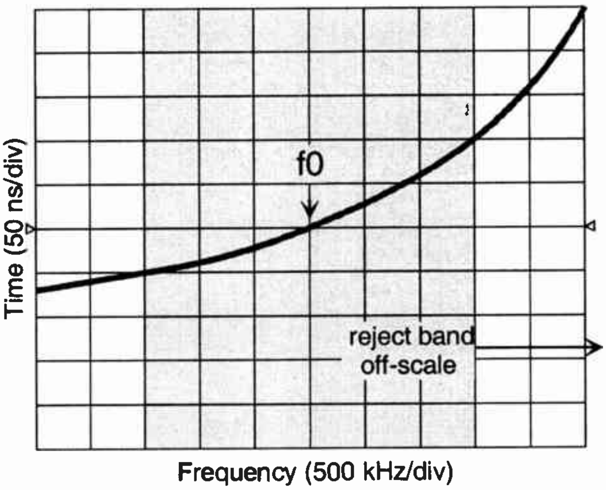


Figure 4.13-26. Group delay for a single notch cavity.

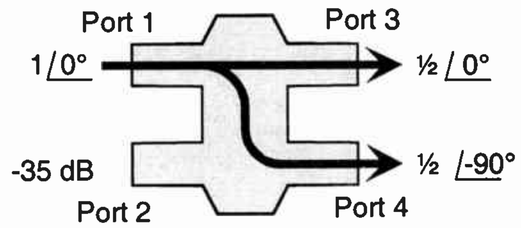


Figure 4.13-27. Hybrid used as a signal splitter.

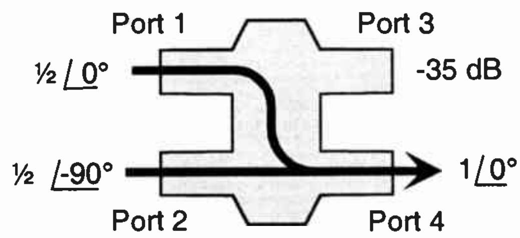


Figure 4.13-28. Hybrid used as a signal combiner.

delay equalization is feasible only for band-pass filtering.

- Expandable site. Because of its ability to attenuate all frequencies except its own, a site with band-pass filtering is easier to expand to include more stations than is a notch filtered site. In a notch-filtered site, if a new station were added, a new notch filter would have to be added to *each* of the existing stations, making their existing notch filter systems even more narrow band and complex.
- Round-trip attenuation. In a band-pass combiner, if enough energy gets by the filter to generate a spur in the transmitter, the spur must then go back through the filter to get to the antenna, and is therefore attenuated a second time. With a notch cavity, if a spur is generated, the notch filter, not tuned for the spur frequency, will not attenuate the spur, and the signal is more likely to violate FCC rules.

Hybrids

A *quadrature hybrid* is a complex broadband device that has the ability to operate in various modes either singly or simultaneously. The detailed electrical explanation of a hybrid is beyond the scope of this chapter.[2] Hybrids are discussed here only with regards to combining systems.

Figures 4.13-27 and 4.13-28 show a typical hybrid. In Figure 4.13-27, the hybrid is acting as a power and phase splitter. When an RF signal is applied to port 1, the hybrid splits the signal, and the phase of port 4's output is delayed with respect to port 3's output by 90°. Port 2 is called the *isolated port* and is usually terminated with a 50 Ω resistive load equal to the characteristic impedance of the device (usually 50 or 75 Ω).

Use of a hybrid in reverse, for combining, is shown in Figure 4.13-28. If two equal RF signals, with the proper phasing, are introduced at ports 1 and 2, the combined signal exits the hybrid through port 4. If the phase of the two input signals is reversed, the signal will exit the hybrid through port 3. Again the isolated port is usually terminated with a matched resistive load.

The hybrid's third mode of operation is called the *reflected mode* and is shown in Figure 4.13-29. When two identical devices with high impedance, such as band-pass filters tuned to another frequency, or band-reject filters tuned to the incoming frequency, are attached to ports 3 and 4 of the hybrid, the signal entering at port 1 is reflected and exits the hybrid through port 2. The characteristics of this third mode make the hybrid useful in conjunction with other hybrids and cavities in filter systems.

Two hybrids can be connected together to form a *hybrid ring*. As shown in Figure 4.13-30, when an RF signal is applied to port 1, the energy flows diagonally across and exits port 4 with very little loss; ports 2 and 3 are isolated by a factor of 35 dB. If the two

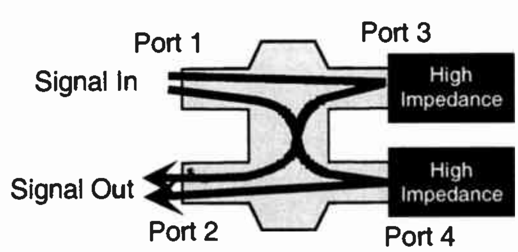


Figure 4.13-29. Hybrid used as a signal reflector.

[2] See Liao, Samuel Y. *Microwave Devices and Circuits*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980. (Chapter 4) *or* Collin, Robert E. *Foundations for Microwave Engineering*. McGraw-Hill Book Company, New York, 1966. (Chapter 6)

Figure 4.13-30. Hybrid ring, showing low-loss diagonal energy flow.

hybrids are tuned for a specific frequency this isolation can be optimized to 40 dB.

The multiple operational modes of the quadrature hybrid make it the backbone of the runout and balanced combiners.

TYPES OF COMBINERS

In the early days of FM combining systems, the easiest way known to combine two frequencies was to use a hybrid ring where one leg contained an extended path equaling an even multiple of both signals' wavelengths. This was accomplished by adding a long length of coaxial transmission line (the runout) between the two hybrids to connect two of the internal ports, as shown in Figure 4.13-31, while the other two internal ports were directly connected. The transmitter-to-transmitter isolation for this combining system depends on the accuracy of the runout line length and the tuning of the hybrid ring.

In a two-station system, this is fairly straightforward. For example, to combine 92.5 and 94.5 MHz, the runout line is approximately 450 feet long.

But for each additional station, an additional combining stage and another long length of transmission line are necessary. The line length for each stage is calculated from the new frequency and the center frequency of the preceding stage. Disadvantage of this method include:

- The runout lines tend to be hundreds of feet long and are difficult to store and protect
- Since each combining stage is tuned to the center output frequency of the previous combining stage, optimization is lost at each stage
- This type of combiner is narrowband, because it is line-length dependent

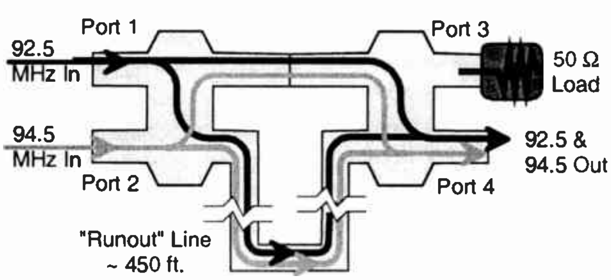


Figure 4.13-31. Runout combiner schematic.

- The isolation is not high enough for today's transmitters. Therefore, filters have had to be added on the transmitter side of the combiner to retrofit existing systems.

A variation of the FM runout combiner used a filter in place of the long runout line to provide the same electrical delay as the transmission line. However, this modification did not substantially improve the electrical performance of the runout combiner, and additional filters have also been retrofitted to these systems.

Runout combiners have almost all been either retrofitted or replaced by more modern combining systems.

Branched or Star Point Combiners

A branched combiner is a simple combination of a tee junction and the required number of filters to ensure a sufficient amount of isolation. For example, an FM branched combiner containing a three-cavity band-pass filter in series with a two-cavity band-reject filter may be used to provide the high quality isolation required for two close spaced (0.8 MHz apart) frequencies (see Figure 4.13-33).

A five-cavity filter system is tuned for one of the channels in a two-station branched combiner, and forms one leg of the combiner. The second leg is a similar combination of filters tuned for the other frequency, and its response plot resembles a mirror image of Figure 4.13-32. The basic schematic, shows the combiner configuration (see Figure 4.13-33).

TX1 and TX2 are the signals from transmitters 1 and 2 as they enter the combiner. The length of the coaxial line between each set of filters and the tee junction is adjusted to provide a very high impedance (approaching an open circuit) to the other frequency, so that the power flow of each signal is through its own filter, out of the tee junction, and up to the antenna.

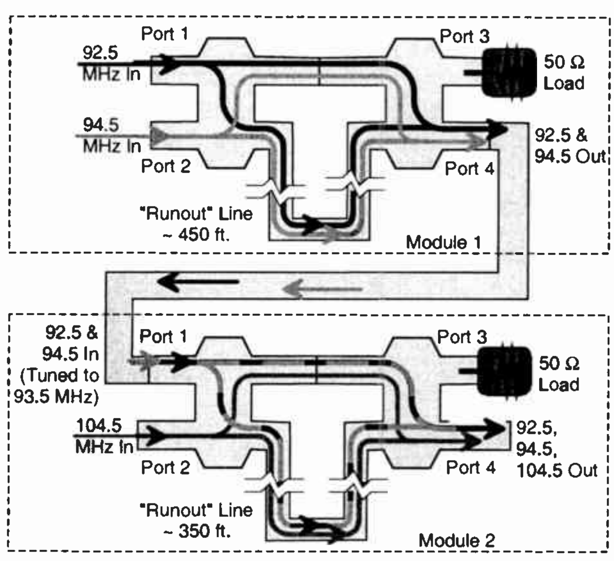


Figure 4.13-32. 3-station runout combiner combining 92.5 MHz, 94.5 MHz and 104.5 MHz.

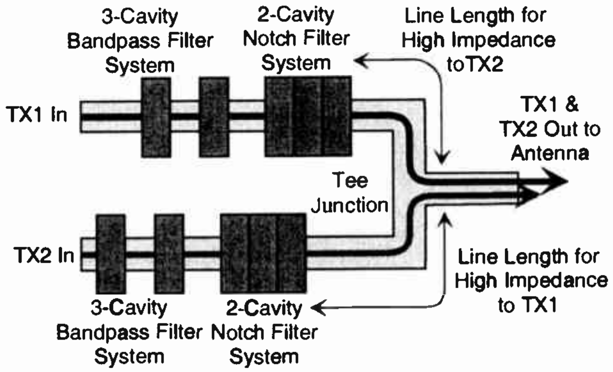


Figure 4.13-33. Branched combiner schematic.

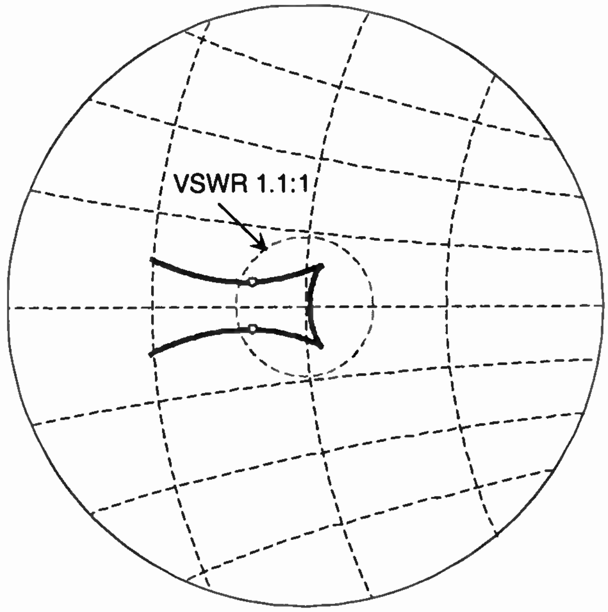


Figure 4.13-35. Impedance diagram, three-cavity band-pass filter and two-cavity band reject filter in series.

Refer again to Figure 4.13-14, curve B (response curve for a three-cavity band-pass filter) and Figure 4.13-25 (response curve for a two-cavity staggered-frequency band-reject filter). When these filters are used in combination, the resulting response from this five cavity system curve is shown in Figure 4.13-34.

Note that the insertion loss for the pass frequency f_1 is only about 0.25 dB and the isolation at the reject frequency f_2 is greater than 50 dB across the channel.

The impedance plot is likewise the combination of impedance plots for the same filter combination (see Figure 4.13-35).

Limitations

A branched combiner is very efficient for a two-station installation, and has been used for as many as four stations, but a tee junction for more stations than that starts to become impractically large, and adjusting the lengths of interconnecting coax becomes more complex. Also, a simple branched combiner cannot be expanded later to include more stations, although it can be expanded by integrating it with balanced combiner modules. To combine more than four stations, a balanced combiner becomes more practical and cost effective.

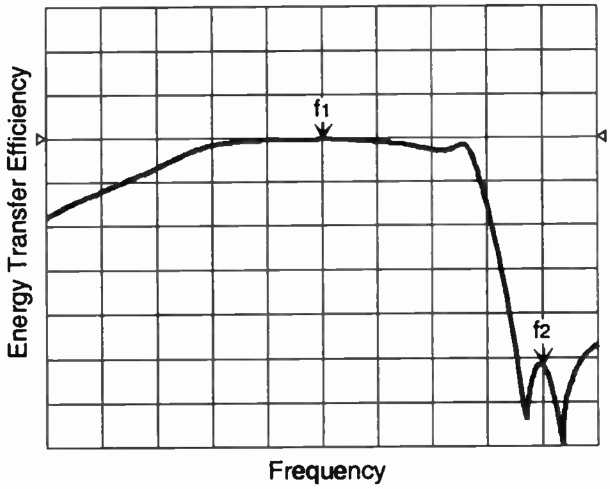


Figure 4.13-34. Frequency response, three-cavity band-pass filter and two-cavity band-reject filter in series.

Balanced Combiners

The balanced combiner, like the runout combiner, is based on a hybrid ring. Each leg of the ring contains an identical set of either band-pass or band-reject filters. Most of the original balanced combiners used notch filters.

Notch Filter Balanced Combiners

In the notch filter balanced combiner (Figure 4.13-36), both notch filters within the hybrid ring are tuned to reflect TX1's frequency, which enters the combiner (black arrows) at port 1. That signal is reflected by the filters, and exits at port 2, utilizing the reflected mode of hybrid operation as shown in Figure 4.13-29.

TX2 (gray arrows) enters the broadband input port of the module, port 3, passing through in the diagonal

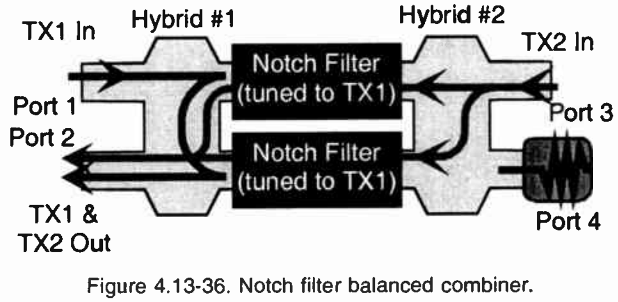


Figure 4.13-36. Notch filter balanced combiner.

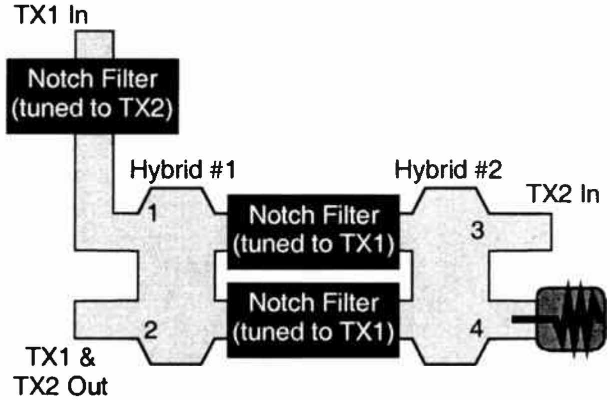


Figure 4.13-37. Balanced combiner with additional filter at TX1 input.

mode shown previously in Figure 4.13-30, with minimal loss in the reject cavities.

The isolation of transmitter 2 from frequency TX1 is the sum of the hybrid ring isolation of 35 dB and the isolation of the notch cavities, which can approach 35 to 40 dB. However, the isolation of transmitter 1 from frequency TX2 is only that of the hybrid ring about 35 dB. Therefore, additional filtering tuned to TX2 is required to ensure that no spurs are generated within transmitter 1. This added filter is shown in Figure 4.13-37.

More modules must be added as the number of stations entering the system is increased. Figure 4.13-37 shows a three-station, two-module combiner. Note the filter in the TX1 input line, as previously described, Module 1's output becomes the input at module 2's port 1. Since both TX1 and TX2 are entering module 2 at port 1, the filter at the TX3 input at module 2's port 3 must now reject both frequencies TX1 and TX2. This can be done with two reject filters, as shown.

External Band-Pass Filtering

A better way to reject multiple unwanted frequencies, of course, is to use a band-pass filter tuned to the desired frequency. For example, in Figure 4.13-38, instead of two notch filters at the input of TX3, use a single band-pass filter, tuned to frequency TX3.

A further extension of this pattern is shown in Figure 4.13-39, showing a five-station, four-module combiner. In this example, each input filter is a band-pass filter tuned to the frequency of that input. If reject filters were to be used at the various inputs, each input would have to filter *all* the frequencies previously introduced. Therefore, port 3 of module 2 would have to contain two notch filters; port 1 of module 3, three filters; and port 3 of module 4, four filters. This proliferation is avoided by the use of external band-pass filters.

Emergency Input Port

In some cases, instead of having a station located at port 3 of module 1, that port is terminated in a 50 Ω load and can be used as an emergency input for any station in the system. Providing an extra port in this way allows a damaged module to be bypassed. Because of the nature of that particular port, as long as the input filter at port 1 of module 1 is a band-pass filter, no further input filtering is necessary.

Limitations of Notch Filter Balanced Combiners

A problem with using notch filters within the hybrid rings is that if the two filters in any one module are

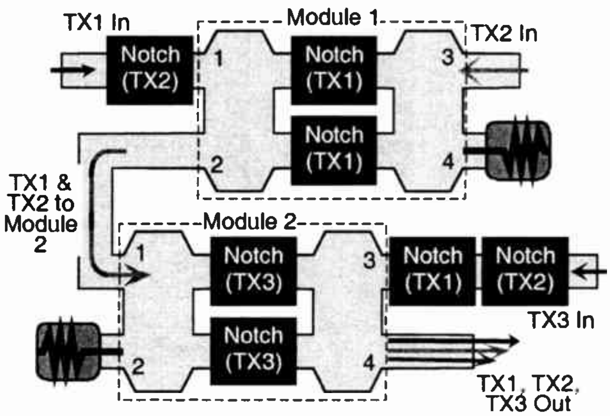


Figure 4.13-38. Three-station, two-module notch-filter balanced combiner.

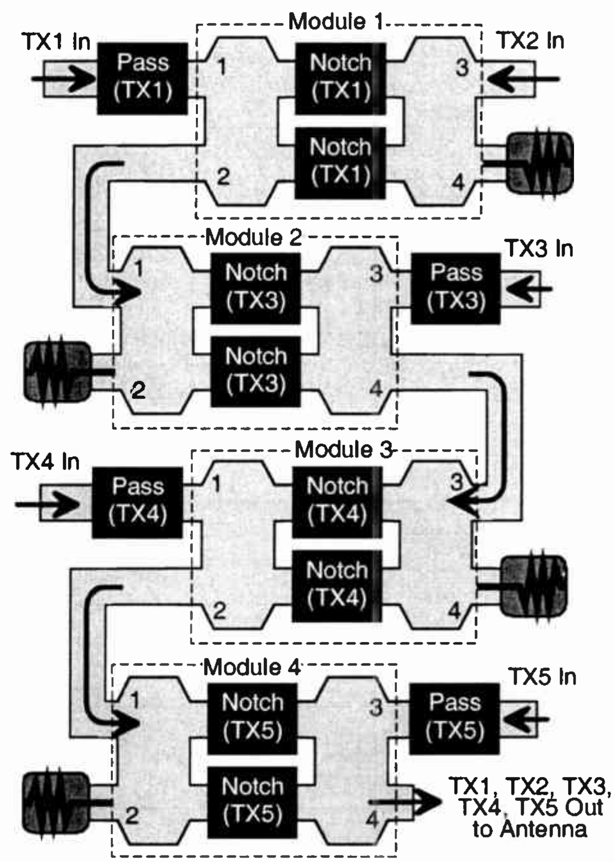


Figure 4.13-39. Five-station, four-module notch filter balanced combiner with band-pass filters at the inputs.

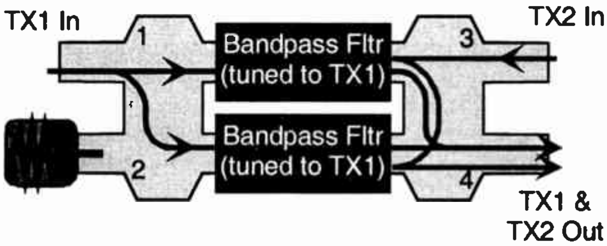


Figure 4.13-40. Band-pass filter balanced combiner.

not identically tuned, an imbalance occurs within the hybrid ring, reducing the isolation to a point where a spur can be generated within a transmitter. Once a spur has been generated, there are no filters within the system to reject that spur, since the notch filters in the system filters are tuned only to the signal center frequencies and not the potential spur frequencies. Therefore, the spur is broadcast.

A second disadvantage of using internal notch filters is that since each module in turn has to conduct the accumulated power of all the previous modules, for a high powered system each module must be larger than the previous one, and the power rating of the system is limited by the size of the final module.

Third, notch filter combiners are impractically narrowband in nature for today's wide channels, especially when the frequencies combined are close spaced.

Because of these limitations of notch filter systems, they are seldom used. Modern FM combiners use band-pass filters.

Band-Pass Filter Balanced Combiners

In a band-pass balanced combiner system, band-pass filters are used within the hybrid ring. The basic system layout is similar to that of a notch combiner.

The power flow is shown in Figure 4.13-40 (compare to Figure 4.13-36). In the notch system, the filters reflected signal TX1 entering port 1. In the band-pass system TX1 also enters port 1, but passes through the hybrid ring's band-pass filters and out port 4, while signal TX2, entering at port 3, is reflected by the filters and exits at port 4.

The isolation of transmitter 1 from frequency TX2 is the sum of the hybrid ring isolation (35 dB) and the isolation of the band-pass filter (about 25 dB).

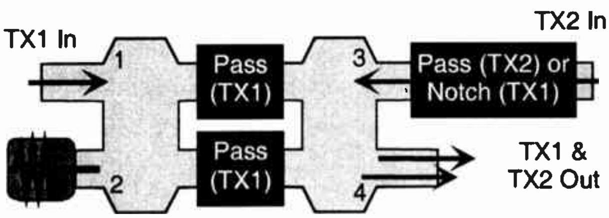


Figure 4.13-41. Band-pass filter balanced combiner with added filter at TX2 input.

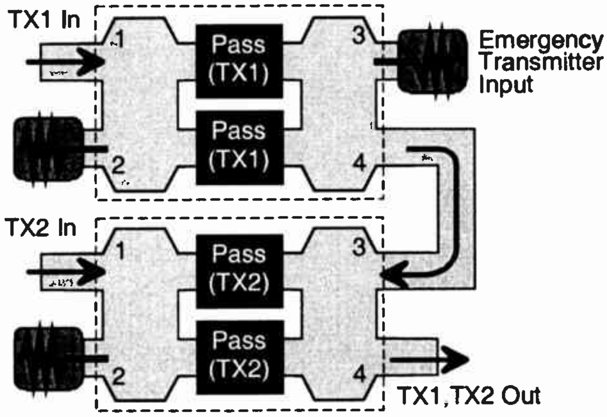


Figure 4.13-42. Two-station, two-module band-pass filter balance combiner.

However, the isolation of transmitter 2 from frequency TX1 is only the hybrid ring isolation of about 35 dB. Therefore, an additional filter must be added between transmitter 2 and its input port (Figure 4.13-41).

Alternatively, a second module may be added to port 4 of module 1, and port 3 terminated in 50 Ω

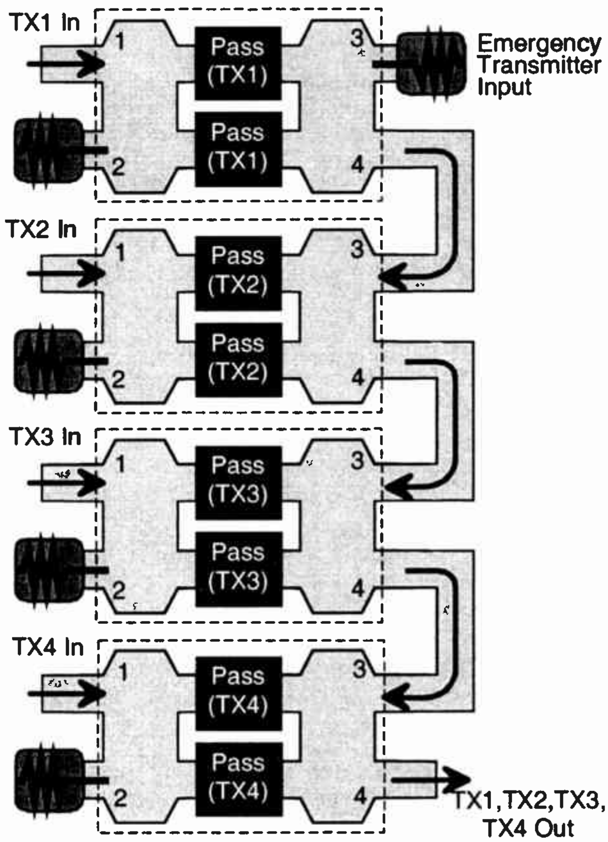


Figure 4.13-43. Four-station, four-module band-pass filter balanced combiner.

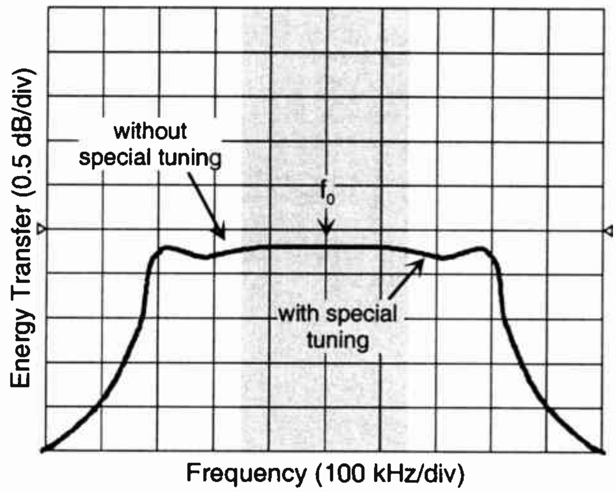


Figure 4.13-44. Frequency response for one station of four-station, four-module combiner.

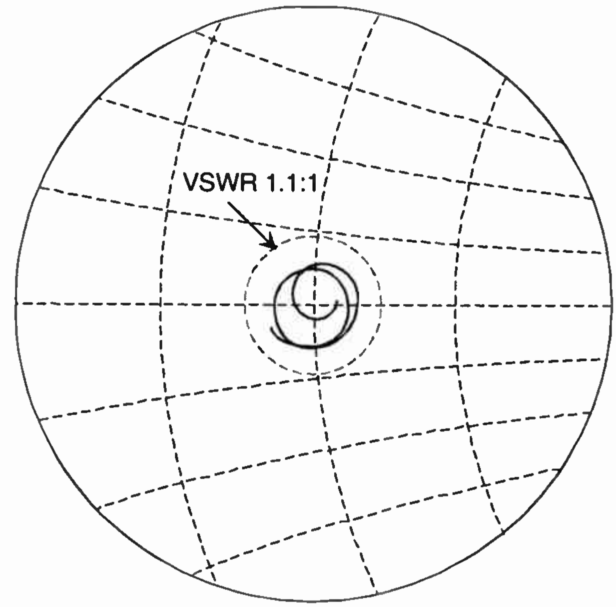


Figure 4.13-46. Impedance diagram for one station of four-station, four-module combiner.

(and available as an emergency input port). Signal TX2 is then introduced at port 1 of module 2 as show in Figure 4.13-42.

No input filter is necessary now for TX2, because it is isolated by the bandpass filters in Module 2. The energy seen at Port 4 of Module 2 is now the combination of TX1 and TX2.

A multiple-station band-pass balanced combiner is an extension of the latter configuration, where each frequency has its own module (see Figure 4.13-43).

In a band-pass system the accumulated power entering each module flows only through the output hybrid, so the power handling capacity of the system is limited only by the size of the output hybrids and interconnecting transmission line, not the entire module.

The frequency response, the group delay, and the impedance diagram for this combiner are shown in Figures 4.13-44, 4.13-45 and 4.13-46, respectively.

Group Delay Effects

When two stations are 1.2 MHz apart or closer, the band-pass filter will not provide quite enough isolation, allowing a small amount of signal interaction. This affects the group delay curve of the module which is furthest from the antenna, as shown in Figure 4.13-47.

This phase distortion can be compensated for by adding a group delay equalizer (see Figure 4.13-48).

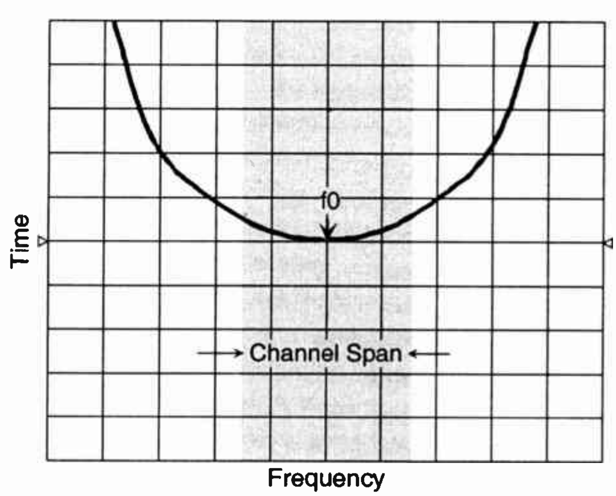


Figure 4.13-45. Group delay for one station of four-station, four-module combiner.

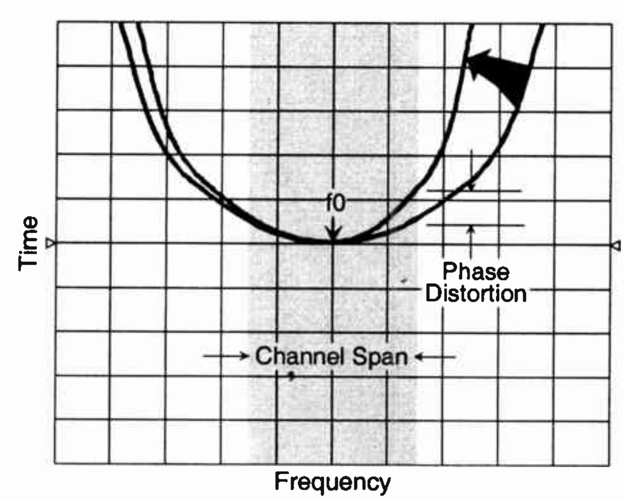


Figure 4.13-47. Group delay diagram showing phase distortion in module farthest from the antenna.

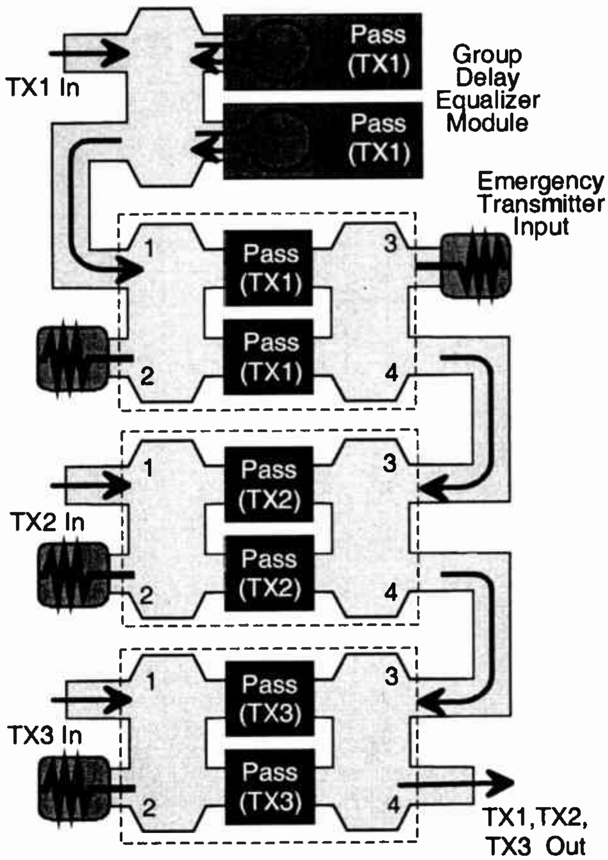


Figure 4.13-48. Band-pass filter balanced combiner with group delay equalizer.

CONCLUSION

Combiners are required when it is necessary to transmit multiple signals from a single antenna. Without proper combining, signals will interact in each other's transmitters, producing intermodulation and spurious products. This section discusses the use of combiners in FM broadcasting. Several fundamental designs and many different components and configurations are described, including some that are now obsolete and are presented mainly for historical interest.

The various types of combiners have their own advantages and disadvantages. The system designer must be aware of each so that he can select the appropriate filter system or systems for his own application.

ACKNOWLEDGEMENTS

Albert G. Friend, technical writer/editor for Shively Labs, edited the text and created the illustrations.

4.14
FM TRANSLATORS AND BOOSTERS

CHIP MORGAN
CMBE, INC., EL DORADO HILLS, CA

INTRODUCTION

Translators and boosters are very interesting from a technical perspective. Although they appear to be easy to design and simple to build, they can actually be more critical than higher-powered stations. In fact, since the signal is local and likely to receive or cause interference, design and optimization methods for these low power systems can make the difference between a successful operation and a failure.

It is a mistake to think that a translator or booster cannot perform as well (within its intended coverage area) as a "full powered" station. There are many low power facilities around the world that provide excellent full-time service to communities every day.

Modern technology provides engineers with the interesting design challenges including synchronized systems, complex three-dimensional directional systems to fill in specific areas, and local systems located in tunnels, buildings or temporary gathering places such as stadiums. These techniques also have application for fill-in service for satellite delivered services as well as terrestrial digital broadcasting.

This chapter does not discuss Federal Communications Commission (FCC) Rules on translators and boosters, since this information is readily available from the FCC. Rather it provides the reader with some ideas and concepts about the successful design and implementation of such a facility.

Definitions

Generally, FM translators and boosters are low powered facilities used to "fill in" the coverage of an FM radio station. Due to the nature of the current FCC Rules regarding service areas of primary stations, it is common for service area contours to extend over areas that actually have terrain shielding from the main signal. If a translator or booster is used to add signal in these shadowed areas, the broadcaster better serves the public within the licensed service area of the station.

FM Translator

A translator is a radio frequency (RF) device that retransmits an FM signal within a specific broadcast band. It alters none of the signal's characteristics except frequency, and its main purpose is to extend or fill in the coverage of a transmitted signal. It typically receives a signal, changes its carrier frequency and then rebroadcasts the signal at substantially higher levels of radiated energy. A translator is essentially a repeater. Listeners within the coverage area of the translator receive primary station programming on *a different frequency from the primary station*. Thus, the station signal is *translated* in that area.

FM Booster

A booster is a RF device that retransmits an FM signal within a specific broadcast band. It alters none of the signal's characteristics including frequency, and its main purpose is to extend or fill in the coverage of a broadcast station. A booster operates on the same frequency as the primary signal and rebroadcasts it at substantially higher levels of radiated energy. A booster is also sometimes called an iso-frequency transmitter. A booster must be located inside the 60 dBuV/m contour of the station it boosts and cannot degrade the original service of the primary station. Listeners receive the programming from the primary station on *the same frequency as the primary station*. Thus, the station signal is *boosted* in that area.

GENERAL DESIGN GUIDELINES

The following tips and points should be considered when designing a translator or booster. Some of these items are simple common sense; others require computer system analysis to compare options and identify predicted results. These guidelines also apply to high-powered systems. In most cases, low powered systems have very stringent requirements due to specific signal challenges. Because the field is quite complex, having a complete understanding of all the technical issues before designing a system will minimize problems. For example, propagation and signal analysis is a career unto itself, but understanding the basics will help you identify potential problems in the initial design of a system. For best results, use the services of a professional who designs translator and/or booster systems regularly.

Initial Needs Analysis

When a low powered station (especially a booster) is being considered to help solve coverage problems in specific local areas, the potential owner should determine the answers to several questions:

- Is there an area with high potential revenues within the service area of the primary station that does not receive signal from the main transmitter?

- Is there sufficient terrain shielding to allow a booster to work technically?
- Does the capital investment required justify the potential improvement to signal problems?
- Do all parties completely understand that when a booster is used, there will be a zone of interference around the booster signal?

If the needs analysis proves value in proceeding, the first step is understanding the process of choosing a location and a system to achieve the goals. Note that the location process outlined below assumes that the reader is familiar with current FCC Rules.

Site Selection

Site selection is the first step in designing a system. The details of the site determine many other design criteria, so an overall review of the impact of any potential site must be quite exhaustive. Choosing the best site for a transmission system is often a series of compromises. Strike a balance between economic, environmental and performance issues of the particular site under review. In addition to these restrictions, the site must satisfy all regulatory, practical and engineering requirements. It must also be accessible during and after construction and in varying weathering conditions (excessive winds, erosion, snow, heat and water).

The primary technical consideration of site selection is the performance of the transmission system located there. Computer analysis should be performed to predict the performance at all intended receiving locations, to predict the interference generated to other operators from the proposed site and to demonstrate the relative quality of any particular site in terms of others being considered. In addition to the performance of the transmitter(s) located at the proposed site, receiving capability of the primary signal must be possible. This may mean satellite reception, microwave link reception or off-air reception of the modulating signal.

When using off-air rebroadcast, site selection is very critical. For example, if a booster receives its primary signal from a receiver tuned to the primary FM transmitter, the booster's transmitted signal generally ranges from 90–110 dB stronger than the intended input signal. If feedback contamination is to be limited to less than 30 dB, the transmitting antennas and receiving antennas must have between 120–140 dB of isolation. Most boosters use other methods of signal delivery, and this is less of an issue in modern systems.

Location of Antenna in Relation to Population

Locate the transmitting antenna site as close to the target population as possible. Building penetration and received signal strength decrease rapidly as distance between transmitter and receiver increases. RF energy diminishes with the square of the distance. If the transmitter site is too distant from the target population, system design becomes much more critical and difficult.

Height of Antenna above Population

More antenna height usually means better coverage. However, excessive antenna height leads to unnecessary interference, increased construction and operating costs and decreased reliability. Specify an antenna height just high enough for adequate Fresnel ellipsoid clearance to the target area but not much higher.

Fresnel Ellipsoid Clearance of Signal into Target Market

The antenna signal is not a narrow ray of energy like a laser beam. The signal is a noncoherent three-dimensional beam that gets wider as it leaves the transmitting antenna, just like the beam from a spotlight. For maximum received signal strength, all the energy in the Fresnel ellipsoid must pass from the transmitter to the receiver with no obstacles affecting the path of transmission. The receiving antenna works exactly opposite from the transmitting antenna. It gathers energy from a three-dimensional area and narrows it down to the location of the receiver, just like a telescope gathers light from a wide area and focuses it on a small area. At the midpoint of the transmission/reception path, the signal beam is its widest. Therefore, even if you can see the population you want to serve from the transmitter site, the broadcast signal may not be able to reach it due to obstructions below the line-of-sight path.

The area of best signal coverage is within the areas receiving the entire Fresnel ellipsoid. These areas can extend beyond visible line-of-sight due to the refraction of electromagnetic waves in the atmosphere. Use a path clearance of at least 0.6 Fresnel zone to avoid excessive diffraction loss due to path obstructions. Path clearances above 1.3 Fresnel zones can yield multipath propagation and nulls in received levels (see Figure 4.14-1).

SYSTEM DESIGN

The strategy of frequency or channel selection for low power systems, particularly translators, is an art form of itself. According to one common theory, the output frequency of a translator should be as close to the input as possible and ideally an adjacent channel. The reality is that practical and technical issues may preclude this possibility. The majority of quality equipment should be able to operate well on adjacent channels, but there are other more complex issues. For example, operation of a translator on an adjacent frequency usually causes interference to the main station. If the translator is located in the center of a large population, it could actually cause more harm than good.

On the other hand, there is no limit to the frequency separation. Somewhere between the two extremes lies the solution. If maximum quality is the design goal, use at least 2 MHz separation for an FM translator so that filters and traditional engineering solutions can be used in case of installation difficulties.

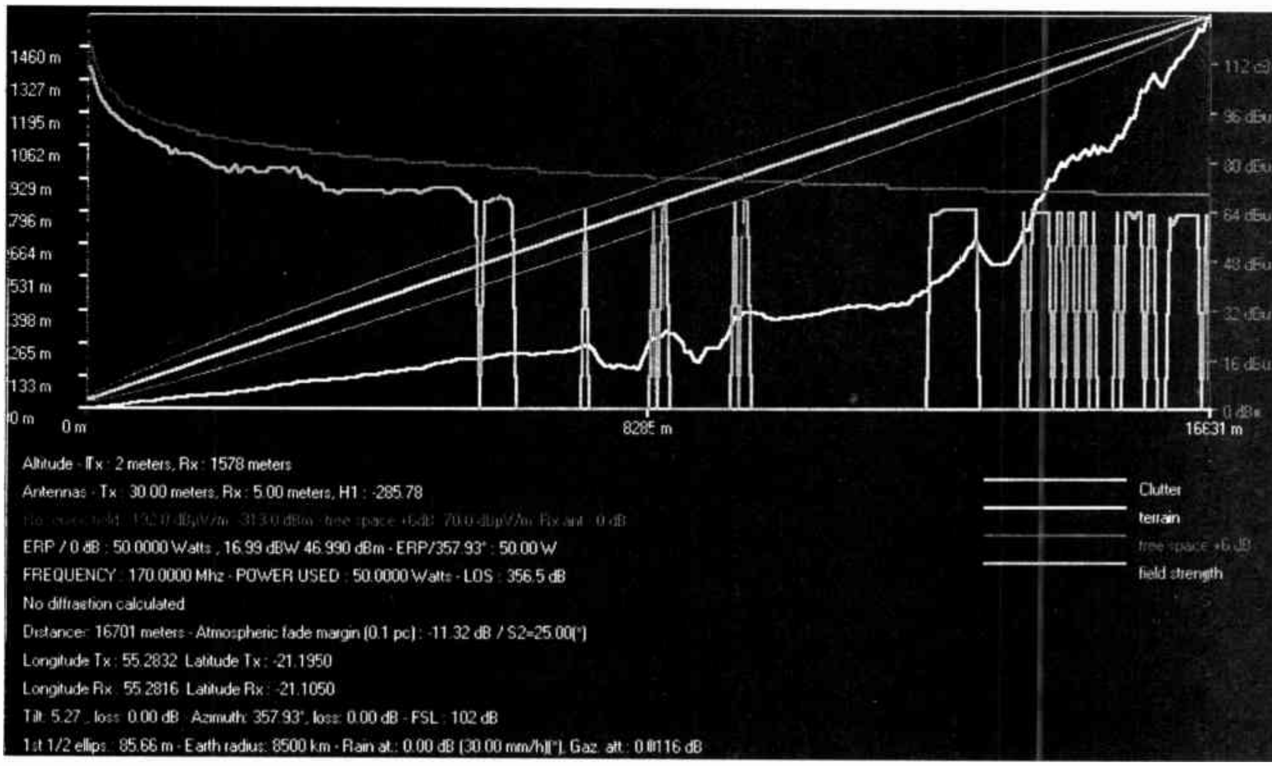


Figure 4.14-1. Fresnel ellipsoid.

Antennas

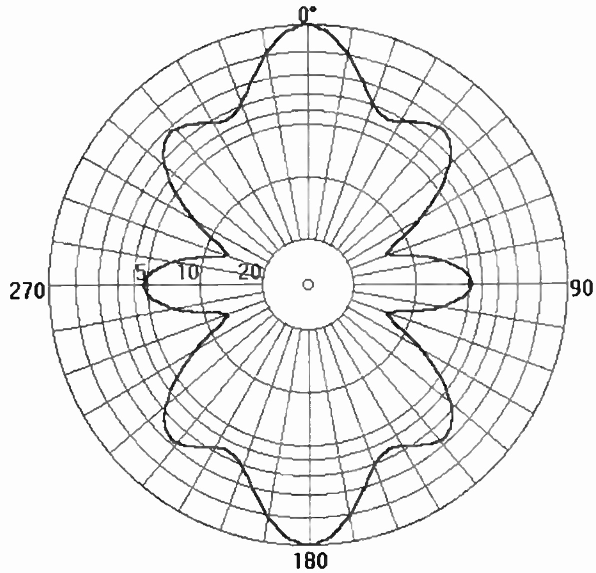
Deciding what kind of antenna to use is a very important aspect of the system. Some system designers mistakenly believe that running a high power transmitter and a low gain antenna is the best course of action in all circumstances. Others believe that a high gain antenna and a low powered transmitter are best. In reality, the best system is designed with specific understanding of the target area and uses the appropriate combination of antenna design and transmitter power to achieve the best three-dimensional pattern beam for the desired performance.

High gain antennas have small vertical beam width that can cause poor reception spots in the minor lobes due to the nulls in the radiation pattern. Since the major lobe provides the most gain to the detriment of areas near the antenna, signals in distant areas will likely be excellent but at the expense of local signal. Sometimes, however, this is an intended effect when lightly populated areas are near the antenna.

If it is necessary to have coverage in both local and distant areas, choose an antenna location high enough to provide adequate illumination of the target area and an antenna configuration that will be the best compromise. A high signal level and a signal path with no obstructions between the transmission site and the receiving sites are required for consistent service to the intended audience. If local coverage is of paramount importance, use a high gain antenna with a moderate degree of beam tilt, or lower power and an omnidirectional antenna. This will make the best use of radiated power and reduce interference. Under all circumstances, analyze the effect the new system will have on co-channel and adjacent channel stations, including the primary station. Antenna pattern control equals better system performance, less co-channel and adjacent channel interference and better spectrum utilization. Early engineers used to say a dime spent on the antenna is worth a dollar spent on the radio, and this holds true even today.

Elevation Pattern of the Antenna

There is often population below the height of a transmitting antenna—especially if it is on a tall tower or hill. The elevation pattern of the antenna identifies the amount of signal radiated toward the horizon as well as the amount radiated above and below the horizon. The elevation pattern below the horizontal plane is critical in transmission systems design because most of the target audience is usually below the horizon. Conversely, a receiving antenna often receives from a location above its horizontal plane, so the elevation pattern above the horizon is important. The number, spacing and type of elements in the antenna array typically control the elevation pattern of an antenna. Elevation pattern gain and beam tilt will not completely solve a problem caused by insufficient Fresnel ellipsoid clearance *shadowing*. If a site has poor performance due to radio shadowing, no amount of manipulation of the antenna system will make a substantial improve-



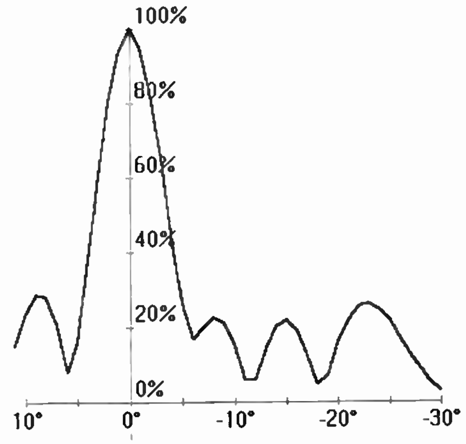


Figure 4.14-2. Elevation pattern.

ment over a properly designed and operating antenna system. Some operators are willing to pay for minor improvements, especially in major markets, but a better use of finances is to carefully analyze a potential site before building it. Examine every proposed system with care to accurately determine the magnitude of the potential problem areas (see Figure 4.14-2).

Antenna Stacking

Antenna stacking increases gain of an antenna array. Stacking which can be either vertical or horizontal applies in transmission and reception and can be helpful in solving certain reception difficulties.

Vertical antenna stacking influences vertical beam width and is effective in amplifying the gain of the array. Horizontal stacking narrows the array's beam width from side to side. If the antennas are in phase, this method adds 3 dB to antenna gain for every doubling of the stack. (A single yagi antenna has a horizontal acceptance angle of approximately 30° between the 3 dB down points off its front). Note that horizontal beam width decreases considerably, and vertical beam width remains unaffected.

A variation in stacking resolves certain interference and reception difficulties. There are a myriad ways to make use of this technique, including offset antenna arrangements, which allow phase reinforcement off the front of the array and phase cancellation off its rear. As an example, set the space from one antenna boom to another so that it will cancel an unwanted signal from a specific forward direction. Install the antennas with the center of each spaced so the unwanted signal is out of phase at the connection between the two antennas, the interfering signals will cancel each other at the combined output (see Figure 4.14-3).

Interference from Adjacent and Co-channel Stations

Consider the potential of interference in the design of a facility. Unless the station is located in an area with very few signals, certain portions of the service area will receive interference from other nearby or more powerful transmitters. This can come from stations as far as three channels above or below the operating channel. Most complaints of poor performance are a result of interference, due to poor signal from the intended transmitter in an area of excessive signal from an interferor or jammer (see Figure 4.14-4).

Translator and Booster Equipment

Once the design is complete selecting equipment for a translator or booster is quite simple. There is a wide choice of equipment available in a range of prices. Of course, the trick is to get the lowest cost equipment while maintaining the desired quality and reliability. To avoid costly mistake use an experienced engineer to assist in equipment choices.

Transmitting Antennas

There are a number of factors to consider when building the antenna system for a low power operation. Careful attention to materials selection is key. A poor quality antenna will adversely affect the best system. For omnidirectional usage, use low power versions of the standard higher-powered transmitting antennas. (They are available in horizontal, vertical and elliptically polarized models.) The relationship between the tower and the radiating elements is essential. The best designs are custom-tuned for the specific situation. Pattern measurements should be performed on an antenna test range, and results of the measurements

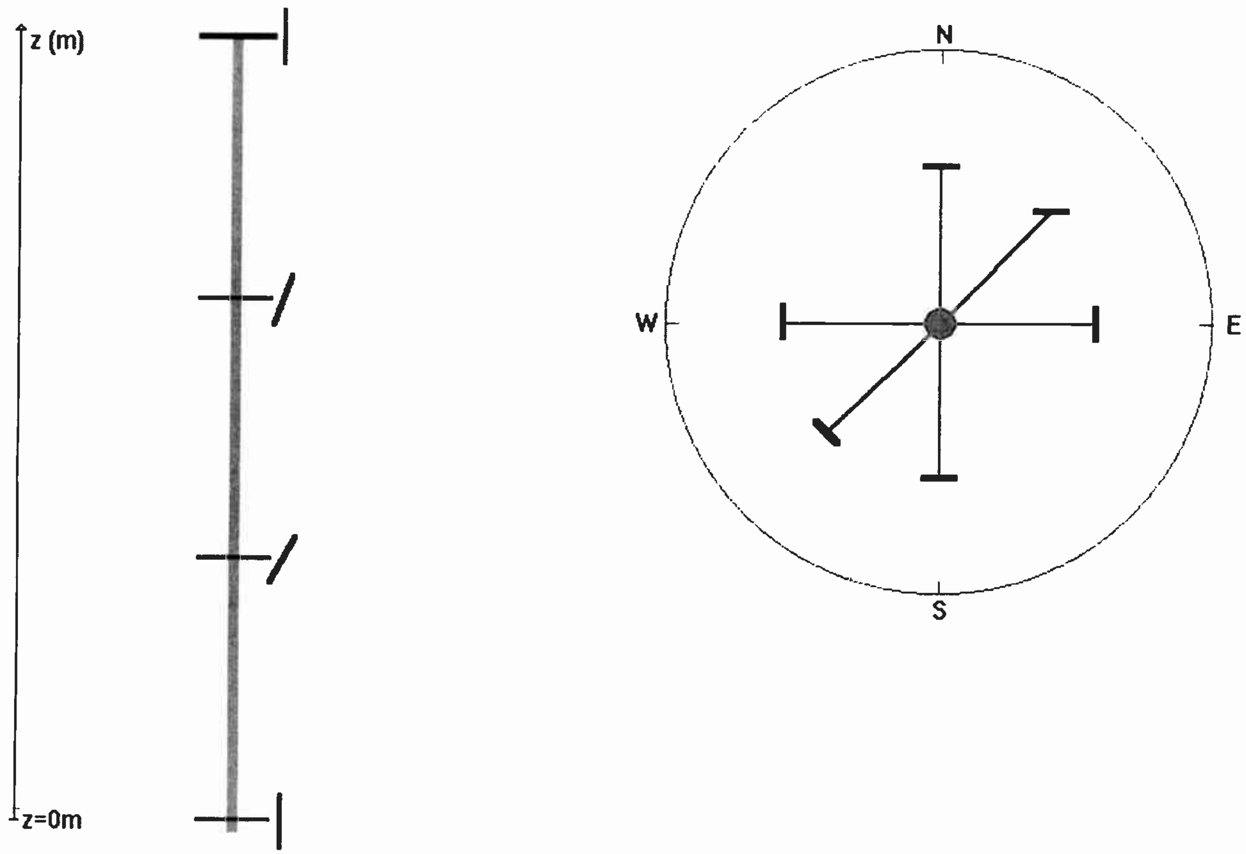


Figure 4.14-3. Antenna stacking.

should be input into a computer generated predictive system to measure the impact of the antenna pattern on the coverage of the system.

When building the antenna system, make sure that antenna and tower manufacturers and the station owners are kept abreast of any possible problem areas such as wind resistance, the antenna's post-installation directivity patterns and mounting procedures.

Receiving Antennas for Retransmission

The input signal's delivery to the transmitter is an essential factor in overall system design when retransmitting a signal. A poor quality input signal affects in the entire operation of the system.

A simple antenna on a nearby tower is sometimes sufficient for off-air reception. More often, more intricate measures are necessary. Sometimes increasing antenna gain or narrowing the pattern of the arrangement is an option. A number of methods can resolve retransmission reception difficulties, including signal filtering to minimize out-of-band products, antenna-mounted pre-amplification to increase signal levels and horizontal and vertical antenna stacking techniques to increase antenna gain and directivity.

The majority of difficulties related to the input signal have to do with weak input signal strength, measured in microvolts on the receiver front panel. Depending on the manufacturer, equipment specifications usually specify that 200–1000 microvolts will result in a good output signal, but those numbers are reflective of equipment capabilities only.

Practical limits at a site tend to be much higher for several reasons. Depending on local conditions, weak signals can result in fades. Upward of 1000–3000 microvolts is usual, but even this level of signal can have problems during a fade. Carefully monitoring of proposed sites helps avoid the problem. The level of RF noise near a receiving antenna can also influence input. Noise levels can be higher than the signal strength of the desired signal. Subsequent filtering or antenna location can help resolve the issue.

In many cases, one must take special measures to ensure that reception and retransmission signals stay clean. A weak signal with poor signal-to-noise ratio (S/N) is generally the problem. The system's receiving section must contain a superior signal conditioning system, pre-amp and receiving antenna to maximize clarity.

Transmitters

Transmitters provide the RF power for the antenna. Proper design of a low power facility includes selection

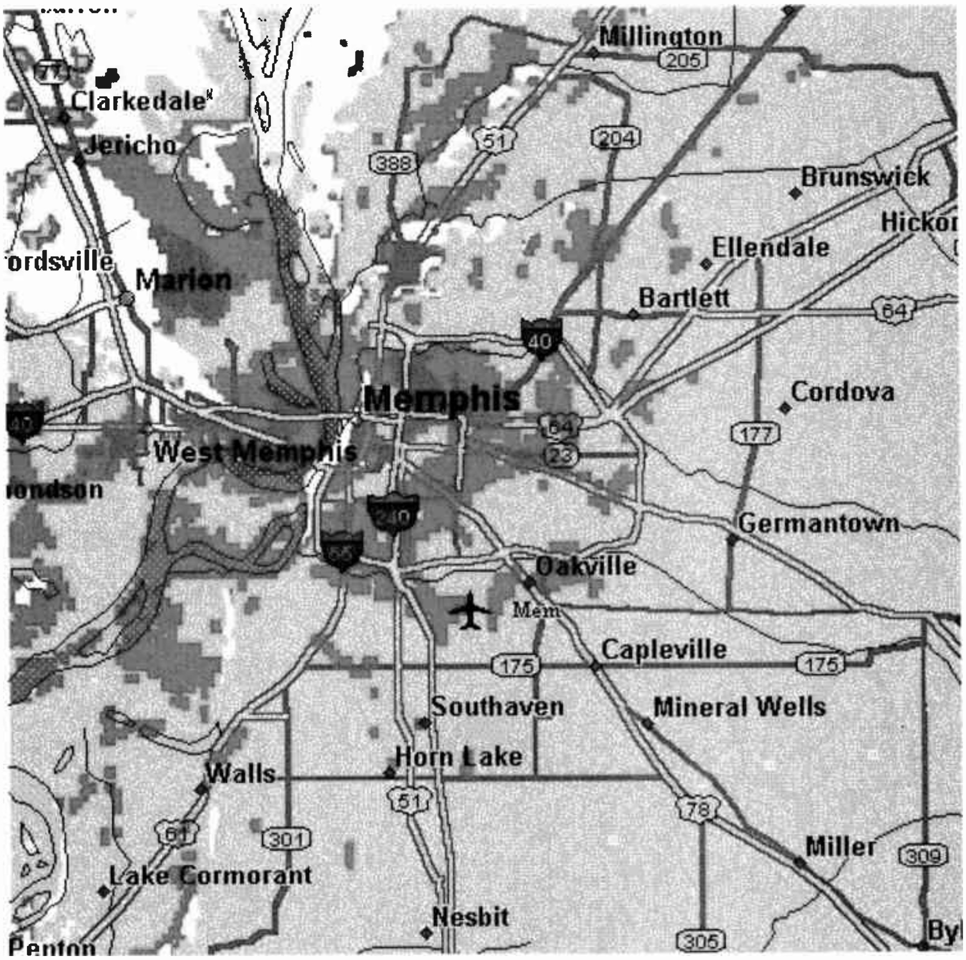


Figure 4.14-4. Interference.

of the appropriate transmitter for the job. A primary consideration is adequate power in order to achieve desired effective radiated power. The gain of the antenna(s) and the length of the transmission line(s) affect this power level requirement as do the available electrical power and the cooling system available at the transmission site. Generally, low power facilities do not need more complex electrical power systems, such as three-phase power or high-voltage power systems. As in any design, the transmitter should be well designed, easy to maintain and able to provide diagnostic information about its status and condition. Low power FM transmitters are usually either a standard FM exciter with low-pass filter or an exciter with a power amplifier. Many of the translators or boosters are solid-state; others operate with vacuum tubes. In a translator or rebroadcast application, several systems are available as complete solutions with built-in receivers and audio processors. In most cases, the transmitter is simple to install and operate.

Transmission Lines

Transmission lines play a major role in the design of a translator or booster. They provide the connection between the antennas and the equipment—whether transmitters or receivers. Detailed engineering data are available to allow the user to accurately calculate and compensate for feedline losses. Use the best cable in low power applications to maximize the power delivered from the transmitter to the antenna or from the antenna to the receiver. The antenna can only radiate the power it receives.

Avoid using a foam type dielectric cable unless it has a rigid outer conductor. Foam dielectric is flexible and has no need for pressurization, but extended exposure to high temperatures can result in migration of center conductors and impedance variations. In some cases, rigid outer conductor foam dielectric cable may be more cost efficient than air-dielectric lines if it is installed carefully without sharp bends or kinks. Keep in mind that air-dielectric cable or some alternate means of delivering nitrogen gas or dry air must be provided if the antenna requires pressurization.

An air-dielectric transmission line has a spiral-wound spacer that runs along its length to hold the center conductor in place. It is harder to handle and install than foam because it is so stiff, but it is much sturdier. Air-dielectric lines have diameters ranging from 1/2–5 in. (1.3–12.7 cm), but the most common sizes for low power operations are 7/8 and 1-5/8 in. (2.2–4.1 cm.)

Here are some important points to remember about coaxial cabling: The coaxial cable and connector quality are just as important as the quality of the system components. All it takes is one bad connector to make a system worthless. Use a semi rigid line for optimum performance. Never use braided cables. Since shield movement can cause noise in systems, secure all flexible cables to keep them safe from strain and unnecessary motion. Become familiar with the dielectric material, as connectors are the weakest link in every coaxial cable system. Avoid nylon connectors—they can absorb moisture as well as high-frequency RF energy. Only buy superior quality connectors from a company with a respected name. Polystyrene and Teflon are excellent choices. Avoid permanently installed adapters. Adapter construction is a compromise providing greater loss and inferior stability than using the correct cable terminations. It can introduce high Voltage Standing Wave Ratio (VSWR) into the system, thereby deteriorating overall performance and decreasing isolation. Above all else, pay attention to manufacturer's instructions. Just about any good quality connector works well if properly installed and protected from cable motion.

Towers

The supporting structure for a low power antenna may be very short or it can be 500 meters tall. Tower design details are beyond the scope of this chapter, but it is important for the translator or booster designer to be aware of various aspects of tower use. Towers are expensive, dangerous and have a tremendous impact on the surrounding environment. Delays in construction can last for years if proper planning and coordination do not occur. In the initial design phase, using of only the best quality tower can help avoid many devastating problems in the end. The loss of a tower can cause more time off the air than the loss of any other major component, it can cause severe damage as well. Pay proper attention to the tower supporting any broadcast antenna. While purchasing the best tower money can buy sounds like an expensive proposition, it is nothing compared to the cost of two towers plus any damage resulting from failure.

Receiving Pre-Amplifiers

Many manufacturers produce tower-mounted preamps with gains ranging from 10 dB to more than 60 dB. Since the received signal may experience increased noise and loss as it travels from the cable to the receiver, mount the pre-amp as close to the antenna as possible. A high quality amplifier should have no effect on signal reception except to increase amplitude to conquer cable and/or system losses as well as render the signal useful. A pre-amp with greater gain does not necessarily mean better performance than one with a smaller gain. The reason is that high gain pre-amps become more subject to overload as signal input increases. Select a pre-amp so that installation and input signal requirements are compatible. Overloading during input will result in distortion that can never be "cleaned up" afterwards. Moreover, if a pre-amp experiences a poor S/N at the input, an equal or slightly worse S/N will be on the output.

Receiver Filtering

There are many ways to implement filtering. In early gain stages of receiving equipment, tuning is broader, and a strong first or adjacent second signal may overload the input or mix in the receiver. Increased sharpness of input tuning can result in an increase in system noise.

Traps or band-pass filters at the input or at the tower may help in certain situations. Usually, a band-pass filter attenuates everything but the desired signal, and a trap may attenuate a specific signal causing interference.

Transmitter Filtering

Band-pass cavities are commonly used to reduce a transmitter's sideband noise. They pass the desired signal with a minimum of loss while sharply attenuating those frequencies that lie above and below the pass-band. Notch cavities have a response curve that is the reverse of a band-pass filter—they possess the same general configuration. Some people call a notch cavity a trap filter or a reject due to its ability to pass the desired frequency while it suppresses the rest. *Pass rejects* (sometimes called *pass notch*) cavities are a combination of the two other cavities. When the interfering frequency is extremely close to the undesired frequency, a *pass reject* is used.

Combiners

Wireless communications usage increases every day. Tower or system space is becoming more difficult to arrange due to environmental pressures, lack of real estate and costs. Transmission sites are becoming crowded. These trends are resulting in a greater need to mix two or more transmitters into one antenna. Combiners do a number of good things at the same time. They filter potential interference and intermodulation, provide isolation between transmitters and allow two or more transmitters to be mixed into one coax run and antenna. Although combiners themselves are relatively, the cost savings to tower sites operators are high.

Cavity combiners are generally manufactured from intercabled band-pass cavities. The isolation is the result of the cavity's resonant response curve. Sometimes, pass reject cavities provide for closer frequency spacing or greater isolation. Advantages of cavity combiners include flexible configuration, cost efficiency and low insertion loss. Negative features include large size, reduced channel capacity and possible redesign of a combiner when frequencies are changed or added. Use a combiner when the frequency spacing between channels is extremely wide. The minimum spacing for the FM band is usually about 1 MHz.

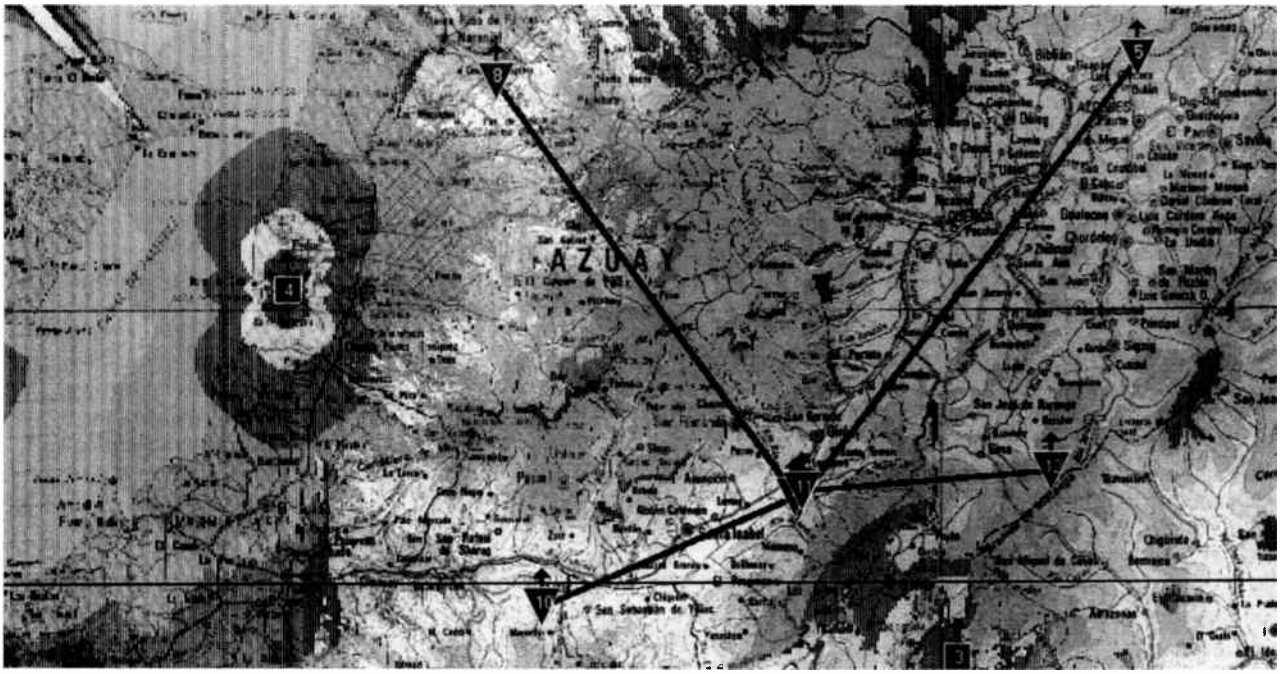


Figure 4.14-5. System coverage.

INSTALLATION AND TROUBLESHOOTING

Every transmitter site has potential for problems. While it is impossible to anticipate every conceivable difficulty, proper foresight and planning can help avoid the likelihood of many disasters.

Tuning and Installing the Antenna

A transmitting antenna is just as sensitive to its immediate environment as a receiving antenna. If you have ever adjusted a rabbit ear antenna on a TV set or the rod antenna on a boombox, you understand the concept. A very small change in the antenna orientation or environment has a drastic effect on its performance. Install the antenna and tune it properly from the beginning, and the facility will perform as predicted. The supporting structure, the antenna design and the way the antenna is mounted all affect the quality of the signal. After installation, the system should be swept with a network analyzer; the coverage should be measured and documented to verify proper installation and for use as a baseline for future review of possible system problems (see Figure 4.14-5).

Adjacent Channel Interference

Adjacent channel interference can be exasperating because the adjacent channel can sound weak and yet be powerful enough to create distortion in the desired signal. When modulation components overlap on adjacent channels, mixing occurs, resulting in out-of-band product generation. The problem can be fixed through special intermediate filters or external filters on the receive antenna.

Condition of Antenna System

Most transmitters have a test meter that indicates if the antenna has failed to the point that it can harm the transmitter. However, there is typically no accurate test equipment available at a transmission facility to qualitatively indicate how well the transmitter or antenna is performing. Most antenna complaints are very subjective. (You just can't "get" the station as you used to. It does not sound right any more, etc.) The only accurate way to tell if an antenna is working correctly is to test it with appropriate test equipment used by antenna manufacturers and consultants. A network analyzer can measure return loss and impedance as well as many other fine details of the antenna system (see Figure 4.14-6).

Quality of Modulation

A commonly overlooked cause of signal problems is poor source material. In this case, no amount of improvement to the antenna or transmission system will correct the problem. Translators and boosters cannot improve the quality of signal they receive. Some systems are implemented with signals from primary stations as much as 100 miles (160 km) away, usually when the elevation is high at the transmit and receive points and low through the transmission path. In such cases, it is important to monitor the received signal for a good length of time using recording equipment, since signal quality can be intermittent.

Spurious Emissions

When there are several stations located at one transmitter site receiving interference, there may be har-

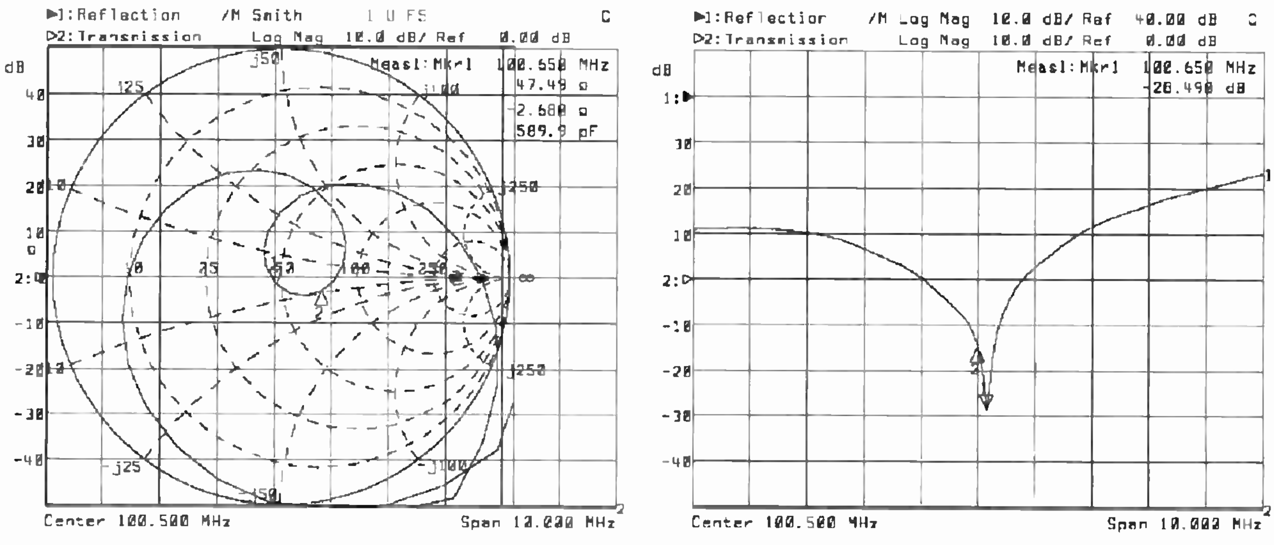


Figure 4.14-6. Network analyzer sweep.

monic mixes between the other stations that fall on or near your frequency. Sometimes these harmonics originate within the transmitters involved in the mix. If the spurious harmonic signals are above a certain threshold, they may be in violation of governmental rules. The existence of this type of interference is verified by using a spectrum analyzer at the towers of the offending stations. The source of this kind of interference can only be verified by connecting the test equipment to the various transmitters involved.

Most complaints are about interference to nearby television receivers. Installing a filter at the output of the TV receive antenna might help. If a filter does not work and the antennas are properly spaced, cabinet radiation or power line feedthrough is the likely culprit.

Serious performance issues can arise when a high-powered nearby FM station's output energy mixes with that of the translator. This is a common problem. Tubes and output transistors are not linear and thus tend to operate as mixers, producing spurious signals that interfere with sensitive receiving equipment. Because the problem goes away when the transmitter is shut off, inexperienced technicians tend to blame the transmitter. The truth is, all transmitters react this way under the same conditions. The goal is to keep external signals from the final RF stage of the transmitter. Installing an isolator in the output transmission line is a cost efficient way to solve the problem. A second option is to employ a high "Q" cavity in the output, which bypasses the transmitter output and rejects the offending signal.

Re-Radiation

Spurious emissions are not necessarily transmitter related. Towers and other metallic objects can re-radiate signals and cause nonlinear distortion resulting in spurious emission. Use a spectrum analyzer to examine the output channel. If the measured signal is clean at the output of all involved transmitters, re-radiated spurs may be the cause. These problems are very difficult to isolate and correct because they are often intermittent. Sometimes introducing mechanical vibrations can help identify trouble locations, but this kind of problem usually takes much time and patience to resolve.

Receiver Induced Third Order Harmonic Interference

When a receiver is near other transmitters, even if the desired station has a strong local signal, the receiver will overload or "be blanketed" by the relatively higher signal levels of the jamming transmitter. The selectivity (ability to reject strong local interference) of the receiver determines its susceptibility to this kind of interference. In either case, harmonic mixes occur within the receiver itself, causing the desired signal to be jammed. This situation is *receiver-induced third-order harmonic interference.* There is no easy cure for this problem except replacing the receiver with a better-designed unit.

Front-End Overload

Front end overload is the result of brute power RF transmissions in the surrounding area. It is entirely possible to have a translator operating on 300 microvolts, for example and a 100 kW station 1 km (.6 miles) away. Front-end overload occurs when the translator tries in vain to accommodate the lower level input signal while being hit by the much stronger signal. Filtering on the input line is the best way to solve the problem before the signals mix in the first place. If the site lacks a strong enough input signal to overcome the insertion loss of a selective filter, consider using special pre-amps designed just for that purpose. Relocating or redesigning the input antenna structure is another viable option. Other methods include raising the receive antenna, adding or stacking receive antennas and relocating receive antennas (or perhaps the entire site). Of course, these methods work for professional fixed receive sites. Vehicles or consumer receivers often have no simple cure.

Feedback

If, after translator installation is complete and the power is turned on, there is nothing but hash and noise, you are probably experiencing feedback. This is because any low power system conveys undesirable products as well as the desired modulation. These products possess a wide bandwidth, which is usually several megahertz from the main carrier frequency. Generated products are comprised of the energy caused by the random motion of electrons (Johnson noise) and sideband components from the desired modulation. Generally, the undesirable products are quite weak, so they cannot be observed on a spectrum analyzer without suppressing the main carrier. Regardless of whether the undesirable products are modulation components or noise, some of these undesirable products will be in phase between the input and the output of the translator or booster. Translator output power is usually in the millions of microvolts while the input signal is just a few hundred microvolts. Poor translator performance is the result of spacing the antennas a few meters apart because the relatively high-powered output puts too much strain on the receive section. A distance of at least 15 vertical meters (49.2 vertical ft) is recommended. To ascertain whether antenna separation is the problem, replace the transmit antenna with a dummy load. Solve the problem with good antenna and frequency separation, usually at least 1 MHz between input and output.

Filters

Providing a filter at the transmitters output can solve many problems. The purpose of the filter is to subdue those undesirable products that fall on a receiver's input channel, but execution is more difficult than it sounds. Note that the quality factor "Q" denotes how lossy the filter will be in operation.

If a translator operates at 100.1 MHz, it would need a filter to prevent undesirable products from falling on the input channel at, for example, 102.1 MHz. Two MHz is the separation factor, so to make sure that no products from the output fall on the input, the bandwidth cannot be more than 1 MHz. $Q = 100.1/1 = 100.1$. This is the minimum value. To have a filter with a low insertion loss and a minimum of energy loss, the Q for every element inside the filter has to be 100 times the minimum, or 10,000 times. However, the realization of such a Q factor necessitates the use of costly cavity type filters.

It is much more cost efficient to use Q factors of around 800 for the filter elements. The filter uses a combination of two techniques: band-stop and band-pass. The band-stop section is a slot about 300 kHz wide, adjusted to the input channel frequency.

The band-pass section passes carrier frequency and useful modulation data. This low loss device adds more selectivity to optimize the system's performance and it can also decrease the radiation of out-of-band signals. A good example would be between an antenna and transmitter combiner, where the filter's job is to overpower any harmonics originating in the ferrite isolators and reduce any broad band transmitter noise so that receive sensitivity remains intact. Energy from the translator output that falls on the input frequency is attenuated by approximately 50 dB. Using an example of 5 microvolts worth of undesirable products falling on the input after passing through the filter, the magnitude is reduced by 50 dB (.0158 microvolts the $s/n = 10/.0158 = 632 = 56$ dB).

4.15
AM AND FM FIELD STRENGTH MEASUREMENTS

**DONALD EVERIST, P.E., COHEN, DIPPELL AND EVERIST, P.C.
WASHINGTON, DC**

AM FIELD STRENGTH MEASUREMENTS

INTRODUCTION

Properly conducted field strength measurements can provide valuable insight on the performance of a broadcast transmission facility. Because of the differing characteristics of AM-band and FM-band propagation, the procedures used for taking AM measurements differ from those used for FM measurements. This chapter will explain both procedures.

AM Field strength measurements are central to establishing and maintaining a directional array. They are performed to determine antenna radiation efficiency and propagation path characteristics. The Requirements for U.S. Broadcasters are given in Table 4.15-1.

Proof-of-Performance

An AM directional antenna is composed of at least two or more radiating elements. The amplitude and phase relations of RF energy fed to each element and the physical spacing between each element produces a predicted directional pattern shape. Each directional pattern is designed and constructed for a specific application, to provide the optimum service to the desired area while providing adequate protection for other co-channel and adjacent-channel stations.

The *proof-of-performance* is used to establish the initial operation of the directional pattern at the time of licensing. It is a condition imposed by the FCC before licensing can occur and is used to update the performance of the antenna system at such times as may be necessary or may be directed by the FCC. It is also used as a reference for subsequent partial proof-of-performance.

TABLE 4.15-1
Selected FCC Rules Applicable to AM Directional Antennas.

Antenna Resistance Measurements	73.54
Antenna Testing During Daytime Hours	73.157
Directional Antenna Monitoring Points	73.158
Directional Antenna System Parameter Tolerance	73.62
Emergency Antennas	73.1680
Equipment Test	73.1610
Operational During Modification of Facilities	73.1615
Operating Power	73.51
Partial Proof-of-Performance	73.154
Proof-of-Performance	73.151
Establishment of Effective Field at 1 km	73.186
Special Temporary Authority	73.1635

Reference Proof-of-Performance

A reference proof is the latest complete proof-of-performance accepted by the FCC, and is the proof to which partial proof measurements must be referenced. The *reference proof-of-performance* specifies, for each monitoring point, the point number, the distance in miles from the transmitter site, the radial and the point location. Section 73.151 of the FCC Rules defines the general field strength measurement requirements to be made on the construction permit and non-construction permit radials for the reference proof.

A reference proof defines the nulls, suppression and major radiation areas of the directional pattern. A reference proof-of-performance requires the taking of non-directional and directional measurements along each radial under similar environmental conditions. While each proof has its own requirements, in general, non-directional measurements begin (as specified by the FCC Rules, Section 73.186) at a distance of five times the height of the non-directional antenna. They are to be made at approximately equal sets of intervals such that, as a goal, 15 to 20 measurements are established within the first 3 km (two miles), eight to ten measurements within the next 7 km, (miles two through six) and 14 measurements to extend to a distance of 24 to 32 km from the transmitter site (miles 6 through 15 to 20). Directional measurements are to be made under similar environmental conditions beginning at ten times the widest spacing between the elements of the antenna system. The instantaneous changing of patterns from non-directional to directional at either a prearranged time or by two-way communications can be beneficial. This technique permits the acquisition of the measurement data in the minimum amount of time and permits the continuous measurement observation without moving the field strength instrument. The accumulation of measurement data whereby both the non-directional and directional measurements are made at a given point at the same time will eliminate relocation and time differences.

Measuring Instrument to Be Used

When field strength measurements are made in support of the partial or reference proof-of-performance, a portable instrument made solely for this purpose and of known accuracy is utilized.

For directional antenna systems, the taking of a proof-of-performance and the availability of the measurement instrument are conditions of the construction permit.

The authorization granted is subject to the following conditions:

> A complete non-directional proof-of-performance, in addition to a complete proof on the directional antenna system shall be submitted before program tests are authorized. The non-directional and directional field strength measurements must be made under similar environmental conditions.

Authorization to Use an AM Directional Antenna System

The directional antenna system requires FCC approval before its operation can commence. Unlike non-directional operations, directional operations cannot be instituted upon a notice and a promise of a subsequent submission to the FCC to fulfill license requirements. The FCC program test authority indicates that the FCC is in essential agreement that the antenna system is in compliance with the FCC Rules and indicates that the permittee has met their basic obligations and authorizes that operation can commence as described in its licensed application. However, specific requests by the Commission may accompany the program test authority and those requests must be satisfied before the FCC will issue a license. The FCC requires a 30 day set of measurements for each directional operation after submission of the proof-of-performance and receipt of the program test authority from the FCC. This request for the data will accompany the issuance of the antenna parameter data and has language similar to the following:

> Program test should be conducted with the directional antenna system adjusted in accordance with the enclosed corrected specifications pending further action on the license application. It is requested that you check the values of field strength at each of the monitoring points at least once weekly during the next 30 days and submit this information in tabulated form to the Commission together with the following meter readings extracted from the transmitter log at the time the monitoring points are checked: (1) common point current, (2) base currents and their calculated ratios, (3) antenna monitor sample current ratios, (4) phase indications, (5) final amplifier voltage, and (6) current.

Partial Proof

Section 73.154 of the FCC Rules governs the taking of a *partial proof-of-performance* as presently required by the FCC Rules. A partial proof must be done, for example, when there is a change of directional operating parameters; when changes are made above the base insulator of the antenna (such as the addition or alteration of an FM antenna or transmission line mounted on the tower), when there is an increase of an existing monitor point license value, or when changes in the environment of the array dictate the necessity of demonstrating compliance with the station's instrument of authorization.

The partial proof-of-performance measurements are to be made at the same locations as specified in the reference proof, but measurements do not have to be taken at all locations. The partial proof-of-performance must contain a minimum of ten measurements at points along each of the radials defined in the reference proof. The partial proof must contain an arithmetic or logarithmic analysis of the measurement data so obtained. It must be demonstrate that the antenna system is operating within its instrument of authorization. Generally speaking, the measurements should be made at an interval of 3 to 16 km (two to ten miles). A statement that the impedance of the common point has been measured and is unchanged from the licensed value prior to the making of the measurements should be provided. A change in common point impedance at the operating frequency requires an appropriate submission to the FCC and is to be requested using FCC Form 302.

The partial proof-of-performance in those areas where ground conductivity is not constant should be made under similar environmental conditions as the reference proof. More accurate information can be obtained when the directional partial proof-of-performance is based upon current non-directional measurements.

General Requirements for Measurements

All measurements should be made during daylight hours in the absence of interference. Special temporary authority may be required prior to the commencement of measurements for a new station. For established stations, the FCC Rules permit considerable flexibility in operation during periods of making field intensity measurements on the antenna system.

FCC policy has been that the measurement observation to be recorded and utilized as a basis of analysis of the inverse distance radiation is that observed with the field set oriented towards the station. The field set maximum indication can have an orientation away from the transmitting source. This phenomenon can be affected by many factors, particularly the depth of a null. Other factors vary from local effects surrounding or adjacent to the measuring point, non-uniform conditions inherent in the propagation path and the position of the monitoring point in region where there is a sharp change in the pattern.

A record must be kept of the measurement data including the point number and description, field strength observation, date and time, pattern under investigation, name of the individual taking the measurements, general weather conditions, and the field strength instrument and date of last calibration. A sample form is provided in Figure 4.15-1 for tabulating the field measurement data.

Monitoring Points

An important, initial step in setting up a directional antenna is the selection of *monitor points* along the

FIELD STRENGTH MEASUREMENTS

FOR____________________

DATE: __/__/____

PAGE NO: ________

ENGINEER: ________

METER: ________

____

N ___ ° E

POINT NO.	POINT DIST.	DATE	TIME	__kW ND MV/M	DATE	TIME	__kW DA-__ MV/M	RATIO DA__/ND	POINT DESCRIPTION

4.15-1. Sample field strength measurement form.

radials specified in the construction permit. These points must be reachable in inclement weather and free of possible reradiation objects such as underground pipes and overhead wires. Schoolyards, churches and cemeteries often provide useful locations since the surrounding area is usually clear.

Prior to submission to the FCC, data should be taken at proposed monitoring points over a period of time until the accuracy and stability of measurements can be established. Readings must be in the direction of the station. This can be established by using a U.S. Geological Survey topographic map or by switching the antenna array to the non-directional mode (if available). See Chapter 1.7 *Broadcast Related Organizations and Information* for information on obtaining USGS maps.

Once the directional array has been adjusted and its performance documented, one of the often overlooked items is the monitoring of the area for new construction. Vertical structure construction or new construction authorized by governmental authorities is often very difficult to detect in the advanced stage of planning. New power line construction and construction of other communication structures can act as reradiators if located sufficiently close to the array. Such reradiated energy can affect the pattern so as to reduce interference protection to other stations.

Graphical Analysis

The inverse distance field or attenuated field intensity at a reference distance (1 km) is that radiation predicted if the earth were to behave as a perfect conductor. As the wave energy travels away from the antenna, this energy reduces in value. The value of the radiated field is reduced by the inverse proportion to the distance from the antenna. For example, if the value of the unattenuated field at 1 km is 100 mV/m, its value at 2 km will be one-half that value or 50 mV/m, and at 10 km, its value will be one-tenth or 10 mV/m.

After the distance from the transmitting antenna to each of the measuring points has been determined and tabulated opposite to the observed field strength values, a measured field strength can be plotted on log-log graph paper. The ordinate is expressed in mV/m and the abscissa is expressed in distance. The FCC, in its conversion to metric, redetermined the frequency curves.

There are 20 frequency charts that encompass the frequencies from 540 kHz to 1700 kHz. The uppermost portion of each chart contains conductivity curves normalized for 100 mV/m at one kilometer from 0.1 km to 50 km and the bottom portion contains the same conductivity curves for distances from 10 km to 5,000 km. (The chart for 670 kHz is shown in Figure 14.5-2.)

For the logarithmic coordinate system, (log-log graph paper) the inverse distance line plots as a straight line. Each of the curves is drawn for the case of an inverse distance field of 100 mV/m at 1 km, and its use is not limited to that value. If an inverse distance field is 200 mV/m (twice the reference value) or 50 mV/m (one-half the reference number) or some other value at 1 km, and if all points on the curve are multiplied by that ratio, this would be the equivalent of moving the curve by that amount on logarithmic coordinate paper. This is the basis by which measurements are analyzed and the appropriate graph for the frequency involved is made by matching the abscissa of the data with that of the FCC graph. By sliding the ordinate information data vertically, the best fit is obtained. By this method, both the unattenuated field at 1 km and the conductivity value along the radial path have been determined. The use of a light table or window or TV set tuned to an unused channel will serve as a backlight for visual analysis of the data.

An individual attempting to analyze measurement data for the first time or not having benefit of supervision can find this is a frustrating experience. One approach is to take log-log graph paper for the appropriate frequency (either the regular or expanded scale) and plot the measurement point values normalized to 100 mV/km. For example, if the non-directional 0.25 kW operation is expected to possess an RMS field at 1 km of 91 mV/m 70° (0.194 of a wavelength) electrical height tower with a normal ground system, see Figure 8, Section 73.190 of the FCC Rules), it has a field 91/100 less than the FCC log-log conductivity graph. Therefore, multiply all values (divide all values if the expected field is greater than 100 mV/m) of the measurement data by the ratio of 100/91 to normalize this data to 100 mV/m. Plot the normalized data. The plotted values can be viewed in relation to the conductivity values if the assumption of the inverse distance field is correct. If the normalized data appears to be over the inverse distance line then the radiation value is higher than assumed and conversely if the normalized data appears abnormally low, the assumed radiation value selected is too high.

This approach can be useful when the non-directional measurements out to 3 km in the various directions have been taken and a quick evaluation of the conductivity values/radiation efficiency around the site is desired. It also will help to assess whether or not the non-directional radiation pattern is being influenced by other adjacent towers in the directional antenna system.

Authorizations

Special temporary authority (STA) will be generally required with any operation which has not received prior authorization from the Federal Communications Commission. Section 73.1635 of the FCC Rules provides that STA be requested in writing (an original and two copies) to the FCC. The authorization request should delineate the station frequency, location, a complete description of the proposed operation and the necessity of the STA. The letter should indicate the capacity of the requesting individual and their contact information so that the FCC can request additional information informally, if required.

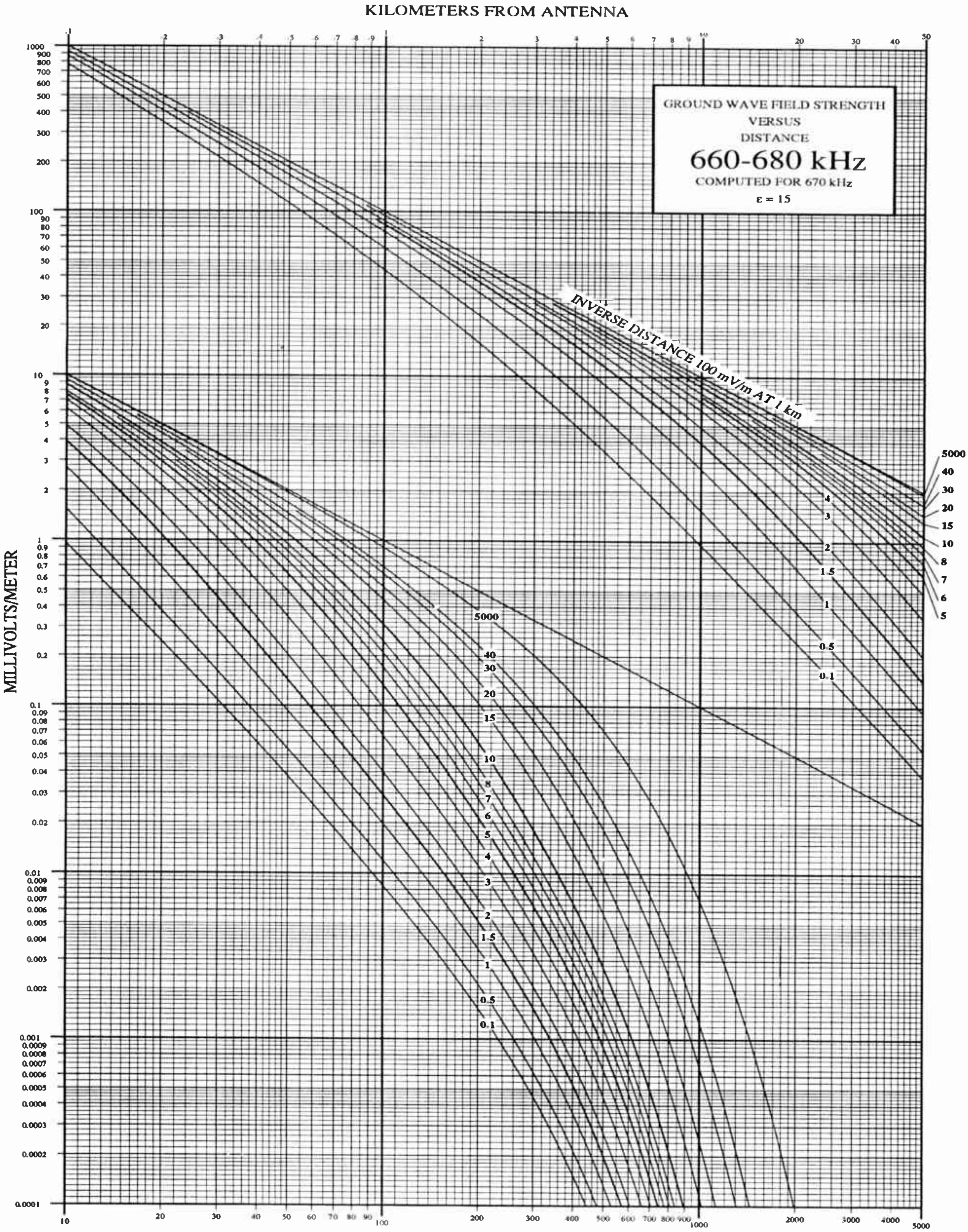
For licensed operations, the FCC Rules provide that non-directional power has been authorized for a proof-of-performance during daytime-only. A station with a single, or more than one, directional pattern for day and night, etc., can utilize non-directional power set forth in the latest proof-of-performance without further authorization from the FCC. However, this privilege is permitted only for the purpose of field strength measurements. In addition, the FCC permits without further authorization the nighttime pattern being operated during daytime hours when field intensity measurements are being taken.

Antenna Monitor System Approval

For each authorization for a new station, details of the antenna monitor system components and installation must be contained in the proof-of-performance in order to obtain the necessary recognition from the FCC that the antenna monitor system conforms to the FCC Rules. An antenna monitor system description accompanying a partial proof must be sent, for prompt consideration, in a separate submission. During construction or revision of the monitoring system, STA for variance of parameters may be required for existing stations. This authority can be obtained by submitting a request (original and two copies) to the FCC. The purpose of the request as well as its duration should be provided. For specific situations, reference should be made to Section 73.68 of the Rules; however, with the revision in Docket 85-90, the FCC is less specific as to the detail of sample-system construction. The FCC indicates as a matter of policy that the methods outlined in the Rules as modified by MM Docket 83-16 would receive continued FCC acceptance. Other less conventional methods may be subject to rigorous scrutiny including observations over a period of time and a partial proof-of-performance. For convenience, Section 73.68 of the FCC Rules is provided as an appendix to this chapter prior to alteration by Docket 85-90.

FCC License Application

The FCC license application is FCC Form 302. Each proof-of-performance must include the information requested in Form 302. It must comply with the provisions and intent of the FCC Rules including Section 73.186 regarding the number of measurement radials, number of non-directional and directional measurements made along each radial, mathematical or graphi-



KILOMETERS FROM ANTENNA
GRAPH 5

4.15-2. Groundwave field strength versus distance frequency curve.

cal analysis, if utilized, plot of the field versus distance measurements on semi-log or log-log graph paper and reproduction of the quadrangle maps (or other maps as required) showing the radials along which measurements were made and locations used in making each of these measurements. Section 73.186 also dictates form and substance concerning the submission of non-directional antenna resistance data as well as the common point impedance measurements with the required graphical plots.

Furthermore, descriptions of the monitoring points complete with the monitor-point photographs and a route diagram as specified by the construction permit must be supplied (see Section 73.158). The station must also comply with other provisions of the construction permit.

A diagram of the RF feed system as constructed (including the phasor, transmission lines, and the tower matching networks) is to be provided. A plot of the inverse distance field at 1 km for the non-directional as well as each directional mode is to be supplied based upon the interpretation of the measured data. The directional pattern must not exceed the authorized pattern in any direction and must have the requisite RMS. In certain situations for new stations or revisions of existing facilities requiring a new reference proof-of-performance, the FCC will permit an adjustment of directional power which can be effected at the time of the license application.

For each directional operation, the parameters as indicated by the antenna monitor for both loop and phase (including sign) as well as the base currents and the ratios are to be furnished. Each field strength instrument and its type number, make and model, and date of the last calibration should be listed. If more than one instrument is utilized, a comparison of the accuracy observed for each instrument should be made.

License

The FCC license provides, among other things, the licensee name, term of the license, station location, main studio location (if not at the transmitter or within the boundaries of the principal community contour) remote control location, transmitter location and its coordinates, type of antenna and ground system (if non-directional) frequency, nominal power, hours of operation and any special conditions. For stations using a directional antenna system, the second page will provide a description of the directional antenna system, the spacing, orientation, and height of the towers and a description of the ground system. Also provided are the theoretical and operating specifications determined from the most recent partial or full proof-of-performance. The description of the monitor points and the maximum limits of the field strength that the points must not exceed are contained on the following pages(s).

Changes and modifications of any of these items require appropriate notification to and approval by the FCC. When receiving a new license, it should be inspected for correctness by comparison with the license application. The operating parameters must be maintained in accordance with Section 73.62 of the FCC Rules. The directional parameters must be maintained with indicated relative amplitude of the antenna base currents and antenna monitor currents within 5% of the values specified in the license, unless other tolerances are specified. In addition, the directional antenna relative phase angles must be maintained within 3° of the values specified in the license unless other tolerances are required.

Monitor-point values must be maintained within the values specified in the license. An increase in an existing monitor point value can only be accomplished by submission of a partial proof-of-performance to the FCC. A change in monitor point location requires submission of a photograph, route diagram, description of the new monitor point and a minimum of ten measurements including the newly-designated monitor point between two and ten miles as shown in the latest reference proof. This information is to be submitted to the FCC.

FM FIELD STRENGTH MEASUREMENTS

FM field strength measurements performed for general determinations of service area require prior planning including safety considerations, study of the topography of the propagation path, proper field strength measurement equipment, documentation and analysis tools.

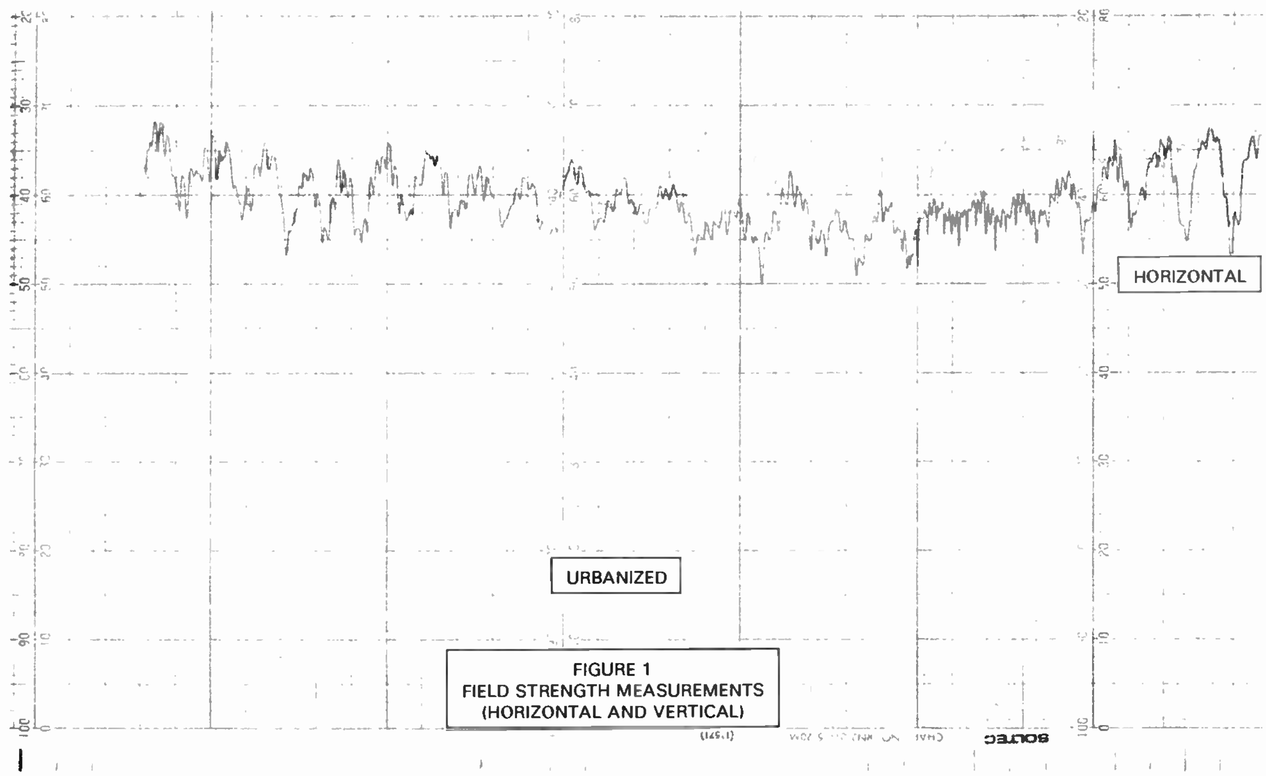
Field Measurements and the FCC Rules

In Section 73.314 of its Rules, the Federal Communications Commission outlines what is required when submitting FM field strength measurements in support of applications. The Commission addresses two methods of measurement—radial and grid. Either of these methods, when properly performed, can be a valuable investigative tool and can yield valuable information about a station's service area on a selective and/or overall basis.

The field measurement data for the station can be helpful, for example, in verifying a site's propagation characteristics to an important service area[1] and in determining an antenna's horizontal and vertical performance characteristic.[2] Performing proper field measurements can be a daunting task since the FM signal level can vary dramatically over a short distance particularly if the measurements are taken with the antenna located close to the ground. The FCC propagation curves F(50,50) and F(50,10) are referenced to a height of 9 meters (30 feet). Figures 4.15-3 through

[1] Alternatively, properly conducted field strength measurements can help define an underserved area. Field strength measurements can provide important information if supplemental service via a booster or translator is being considered.

[2] Field strength measurements can also be made to determine radio frequency levels and spurious emissions.



PERFORMED WITH POTOMAC INSTRUMENT MODEL FIM-71 METER
WITH MANUFACTURER'S SUPPLIED CABLE AND CALIBRATED ANTENNA AT 9.1 METERS

4.15-3. Field strength measurements (horizontal and vertical) urbanized terrain.

4.15-5 provide an example of consecutive horizontal and vertical data recorded at 9 meters in generally urbanized, hilly and flat terrain. While this measurement data is reasonably well behaved, it does provide insight as to the degree of difficulty in attempting to gather reliable and meaningful data under varying propagation paths.[3]

The Commission's Rules provide guidance for the taking of field strength measurements. Briefly, these are:

- field strength measurements are to be taken with an antenna height of 9 meters (30 feet) above ground
- field strength measurement data are to be continuously recorded (graphically and digital) over a path segment of specified length of at least 30 meters (100 feet)
- field strength measurements are to be performed with an antenna and field meter designed to accurately measure the FM signal
- the manufacturer's calibrated antenna is to be oriented for the maximum signal. Each field strength measurement point is to be made in an area which is representative of the area and free of overhead obstructions
- documentation should include the use of a GPS receiver for coordinates, or marking the location on the latest USGS quadrangle plus the date and the time, and the description of the point including topography height and the type of vegetation, buildings, obstacles, weather, etc.
- analysis of the field strength data must be based on the median value of measured signal if it is to be referenced to the FCC F(50,50) propagation curve.

Planning

To minimize time and effort, the purpose for taking the measurements must be determined prior to taking the first measurements. If the inquiry is to be an examination of an aging antenna's radiation performance, this may require field strength measurements to be taken in the eight cardinal directions. Field strength measurements to characterize antenna radiation performance could require the use of a reference antenna with known gain characteristics and effective radiated power (ERP). Also another antenna on the same tower as the antenna under test and whose radiation performance is known, can be used. This could be a top

[3] Measurements were performed with a specially designed vehicle that uses a method in which the chart recorder is dependent on the distance traveled. This permits field strength measurements to be repeated.

mounted FM or VHF television antenna. The mathematical relationship between the desired FM signal and the signal from the reference antenna can be determined and results compared with those anticipated. In any measurement program, measurement points located in clear and unobstructed areas are required. The number of measurements along each radial should be sufficient to establish the antenna's performance in a particular direction.

To investigate an area in which reception difficulties (low signal, multipath or interference) are being experienced will require organization of all factors pertinent to the inquiry. These service area concerns or locations, where possible, should be compiled by zip code and address. Also, additional information such as whether the complaint is from a stationary or mobile receiver, at what time frame and other particular factors must be documented so that they are available when performing the measurements.

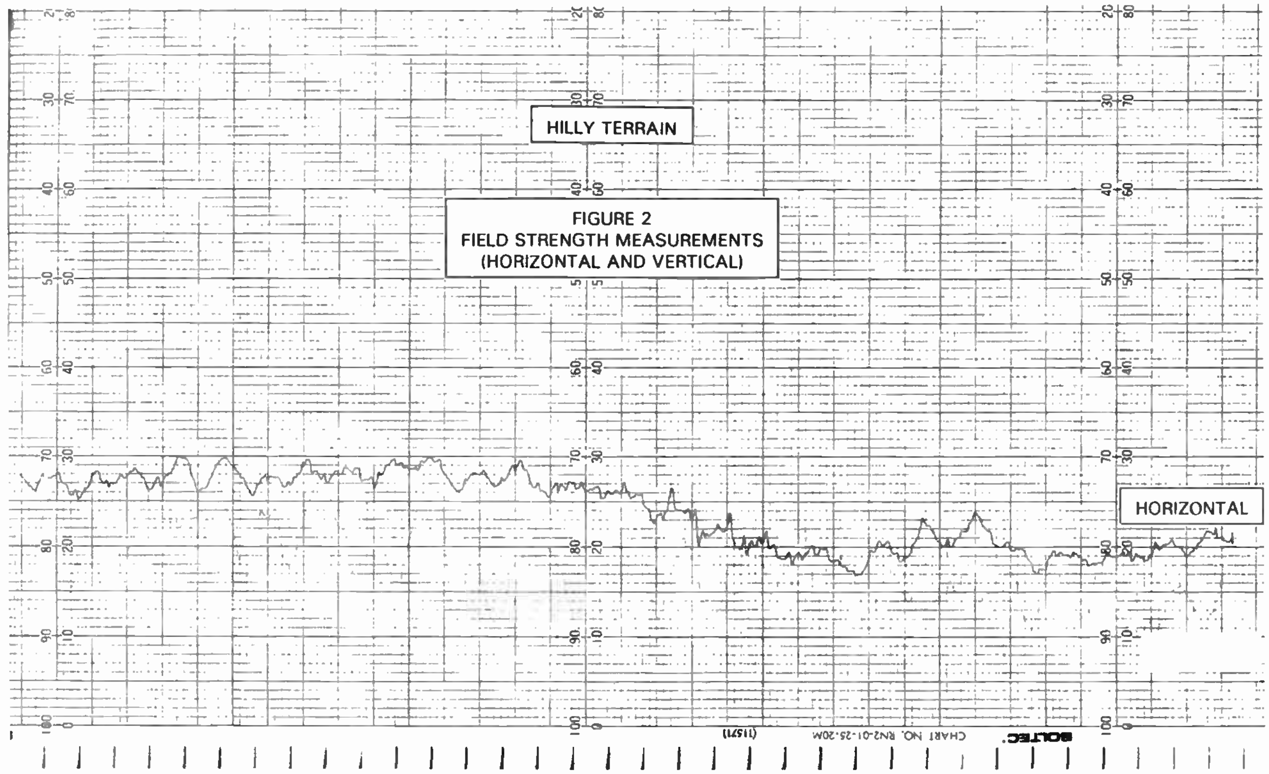
When sufficient information is available to warrant further investigation, several approaches may be used. Initial physical investigation could be performed by documenting the area of concern by using several types of good quality receivers. The results of this initial investigation could be augmented by manually plotting profile information or by using one of a variety of computer profile programs to ascertain the nature of the propagation path.

From this preliminary examination, if the cause of the service deficiency is not apparent, expansion of the study by making field strength measurements may be warranted. The preliminary examination, prior to measurements, must yield sufficient information to give focus to the nature of the further investigation. In essence, complaints by advertisers and the listening public must be verified before taking any measurements.

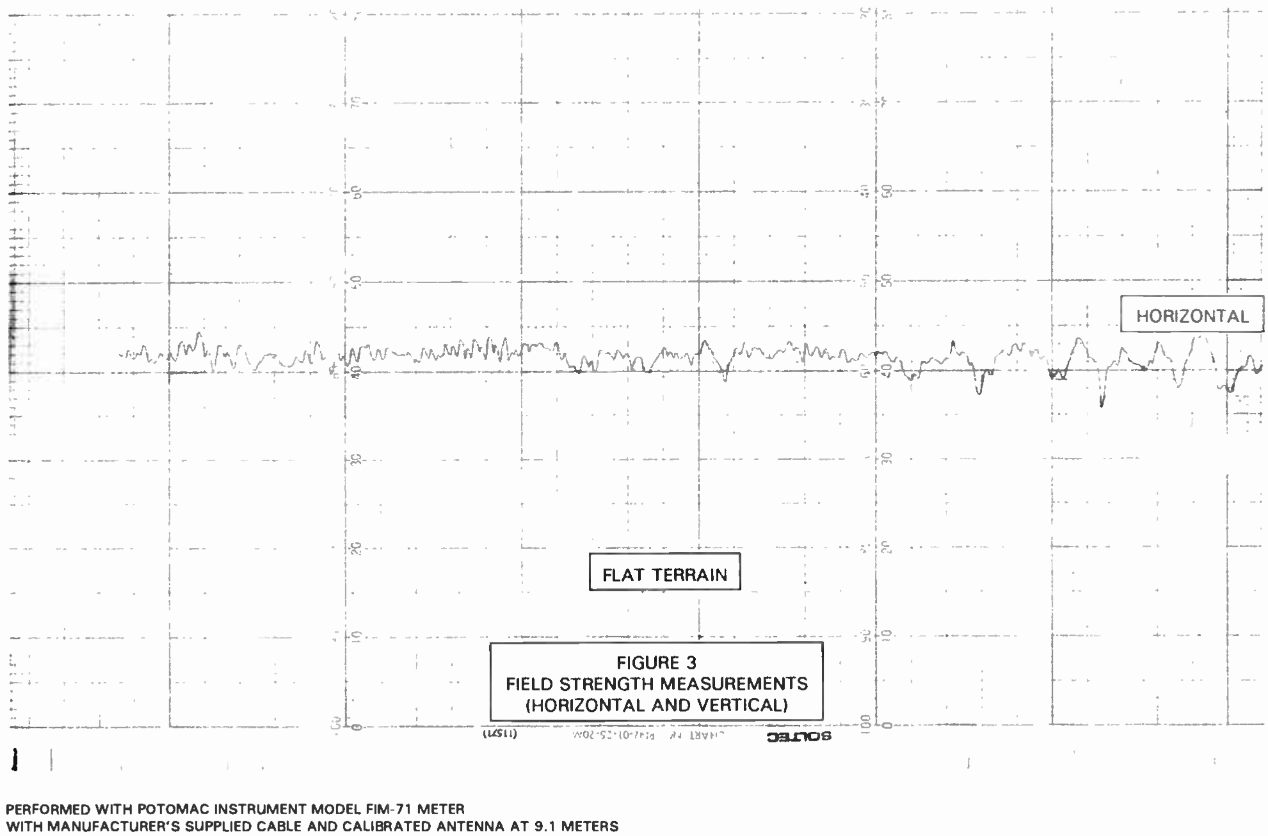
Data Analysis

Data analysis can be performed manually or through the use of computer software. Analysis techniques should be configured to adapt to the measurement program. Mathematical or graphical analysis may be appropriate.

Figure 4.15-6 shows the Commission's F(50,50) propagation chart (from Section 73.333, Figure 1 of the Rules). This chart describes the FM signal levels versus distance at a particular average terrain value. The average terrain value is determined by an evaluation of the terrain from 3 to 16 radial km from the antenna location (see Section 73.313 of the Rules). The curves can be used to extract a set of data to be



4.15-4. Field strength measurements (horizontal and vertical) hilly terrain.



4.15-5. Field strength measurements (horizontal and vertical) flat terrain.

compared to the measurement data. For example, along a particular radial, a value can be determined which best describes the terrain in which the measurements will be performed. The second piece of information is the appropriate ERP in that direction in dB (for example 50 kW = 17 dBk, 100 kW = 20 dBk, etc.). Next, a distance needs to be selected. From these steps, using the F(50,50) propagation chart, a new curve can be generated from a set of values that characterizes propagation path in terms of predicted signal level. These steps are performed for arbitrary distances such as 1 km, 5 km, 10 km, etc. There are illustrations on Figure 4.15-6 which show that, for a 50 kW ERP at a height above average terrain of 100 meters, the predicted signal level at 10 km is 87.5 dBu and for 20 km is 75.4 dBu. This data can be used at specific distances to provide a comparison to the measurement values. Alternatively, a set of points can be developed such that a curve can be plotted if a graphical approach is desired.

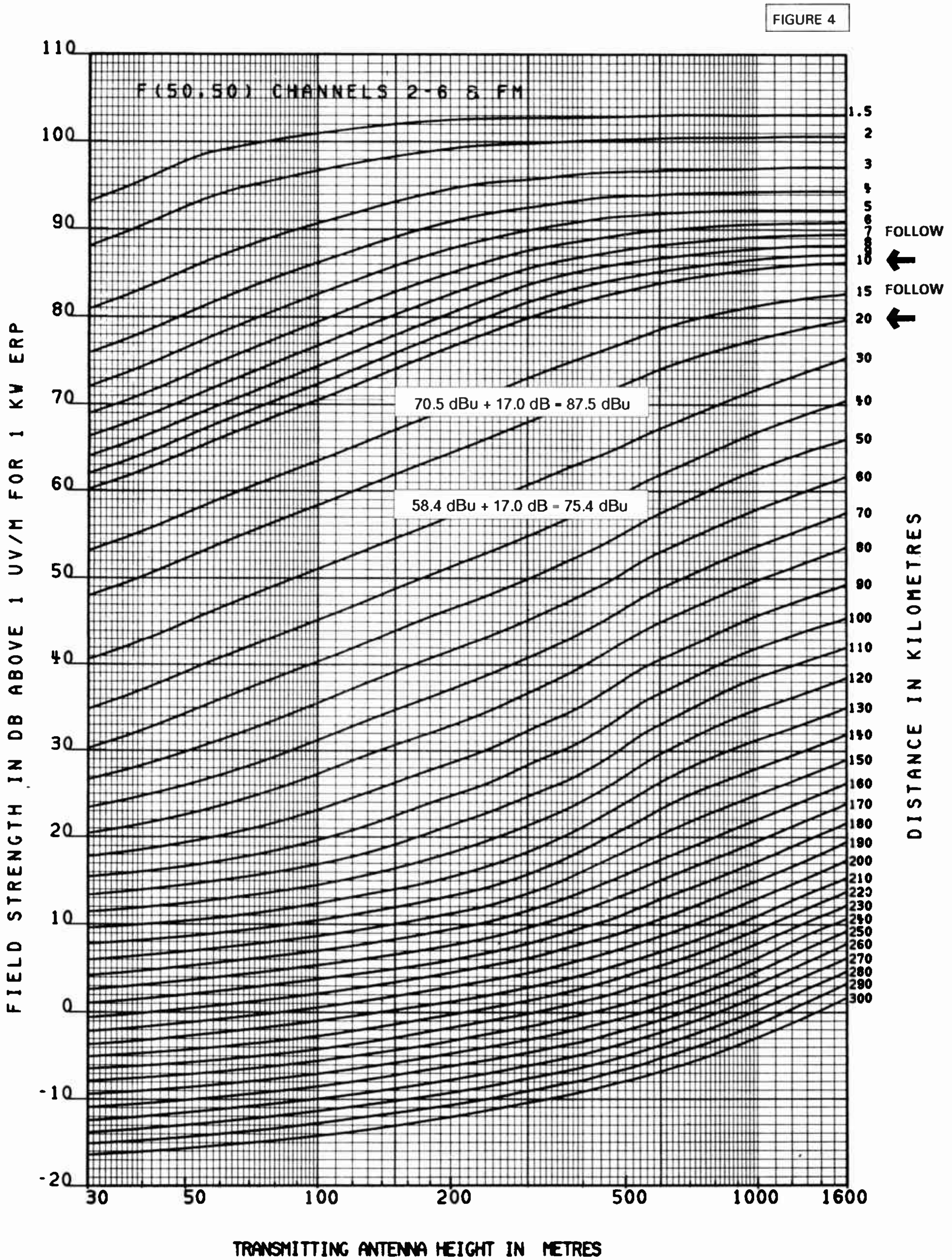
Safety

Performing measurements requires that all safety considerations be made, including every facet of the measurement program. This includes proper mounting of all measurement equipment, raising and lowering of the calibrated antenna, point selection methodology and ensuring that there are no overhead obstructions and navigating the measurement vehicle in traffic.[4] Measurements of this type must never be taken at nighttime. In addition, if field strength measurements are going to be performed for extended periods, then appropriate rest intervals should be implemented so that operator vigilance can be maintained.

Field Measurements

Horizontal and vertical measurements performed at 3.0 meters (10 feet) can be constructive in documenting difficult reception areas. This technique is used to gather general reception information so that other measures of study can be implemented as warranted. For more difficult situations, enhanced measurement techniques using a receive antenna height of 9.1 meters (30 feet) should be used. It is to be emphasized that for a meaningful measurement program, appropriate planning, details of the transmission equipment, measurement equipment geographic maps and analysis tools are required.[5] In low signal reception areas, but close to where other higher powered transmitters are

[4] This office has continuously maintained a measurement vehicle for over 50 years. In the 1950's this office had a partner killed when the antenna inadvertently encountered a power line.



4.15-6. The FCC F(50,50) propagation chart.

located, it may be necessary to employ a filter or a pad at the input of the measurement instrument. This is to minimize the receiver overload in the measurement instrument. It is imperative that a shielded line be used with good quality connectors. In low signal reception areas where intervening terrain may be a factor, an elevation profile study using 4/3 earth radius of the transmission path using 7.5 minute USGS geological quadrangles may be appropriate. For initial studies, profile radials using 4/3 earth radius generated by a computer database may be sufficient. However, the computer generated profile data should be confirmed in critical areas by the use of current USGS maps. By plotting the center of radiation of the transmitting antenna above mean sea level and carefully drawing a straight line to each reception area under scrutiny may give insight if terrain is a factor. From this study buildings or other manmade obstacles can be added. If a clear propagation path is evident, then examination of the transmitting antenna and its placement should be made.

Transmitting Antenna

A well known phenomena occurs when an FM transmitting antenna is side-mounted on a wide-faced tower.[6] Interaction by the FM signal with the tower members will modify the FM antenna radiation pattern. This can be confirmed by making a number of radial measurements at clear and unobstructed points. If this condition occurs, the horizontal and vertical signal behavior will typically be modified differently. The reason is the effect of the individual tower members upon the horizontal or vertical radiated signal is a function of the tower member dimension, position, and its physical relationship to the radiating element.

Field Measurement Configuration

The configuration of vehicle-mounted measurement equipment can be adjusted depending upon the vehicle. It is essential to have all measurement equipment appropriately mounted in a stable configuration and secured. For all measurement programs including those using non-dedicated vehicles, good documentation of the measurement configuration and the equipment is required so that the measurement program can be reestablished at a later date and time.

A simplified method of making spot measurements is provided in Figure 4.15-7. For example, vehicles with a hinge mounted rear tire support can be modified to permit four uniform height measurements at a single vehicle location. Also provided is a sample chart (Figure 4.15-8) which can serve as a guide for documentation.

CONCLUSION

Field strength measurements are a valuable tool for broadcasters who want to determine exactly how their transmission systems are performing, and how their coverage is being affected by such things as their towers, propagation path and other nearby signals. In the case of AM stations, these measurements are generally required as a condition of license. For FM, they are usually an optional tool available to the broadcaster. In either case, understanding the basic procedures for taking these measurements is a necessity for broadcasters performing their own measurements, and for broadcasters who will be hiring an expert to perform the measurements.

APPENDIX A

FCC Policy Statement Entitled, *Criteria for Approval of Sample Systems for Directional AM Broadcast Stations,* Dated December 9,1985

On October 31, 1985, the Commission adopted a *Report and Order in MM Docket 85-90* concerning the antenna sampling systems and proof-of-performance for directional AM broadcast stations. The new rules are based upon performance standards in terms of accuracy and stability rather than upon construction specifications. This Notice clarifies the information required for directional AM sampling system approval under the new provisions of Section 73.68(a) of the Rules. As before, stations constructing new antenna systems pursuant to a construction permit must obtain approval of their sample system by informal request to the FCC in Washington, D.C.

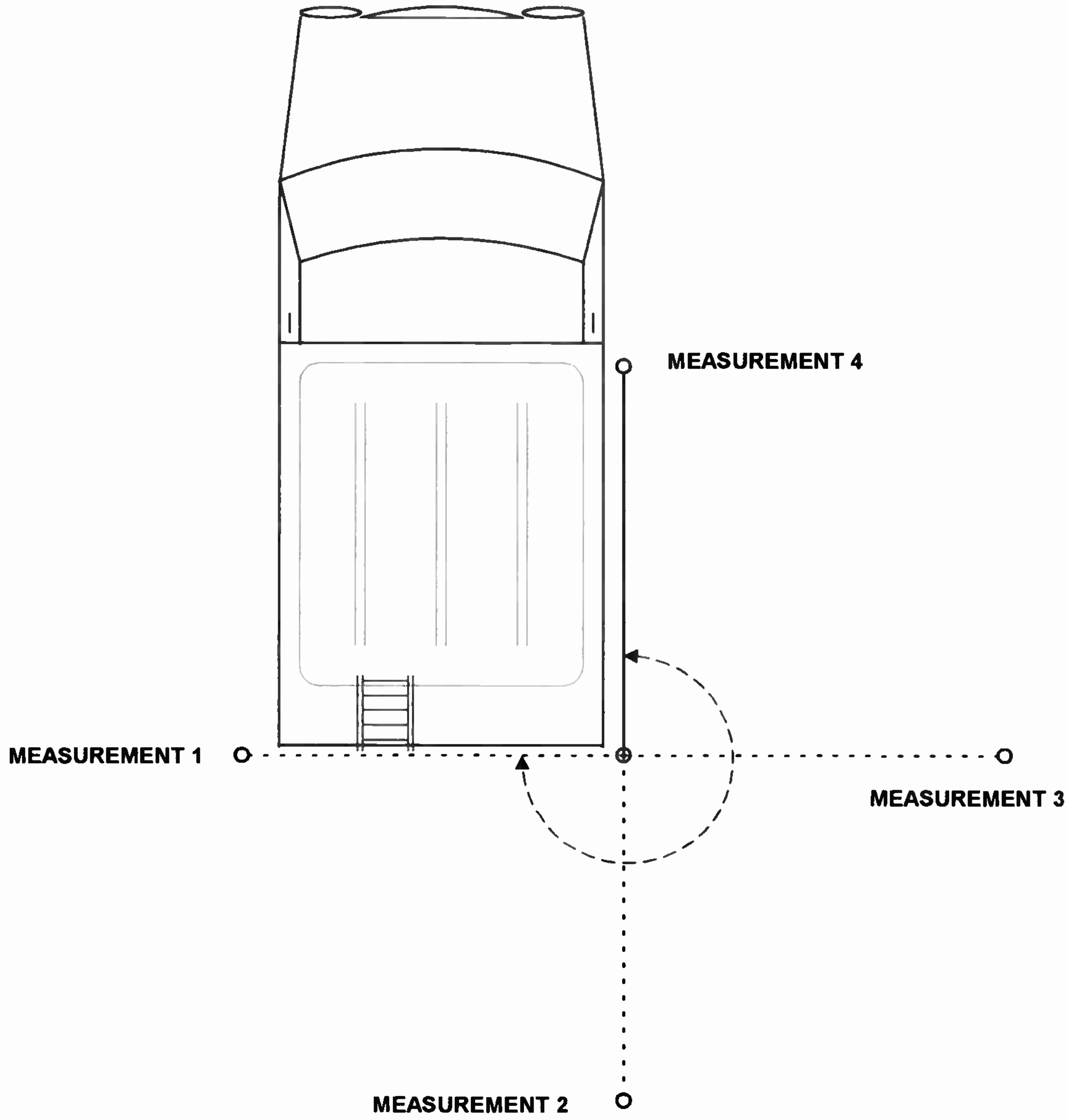
To obtain antenna system approval, applicants may follow either of the procedures set forth in Paragraphs A or B below:

A. Demonstrate that the system complies with the provisions of Section 73.68(a) of the Rules in effect prior to January 1, 1986.
B. Demonstrate stability of operation by submission of the following information:
   1. A detailed and complete description of the antenna monitoring system installation.
   2. Field strength readings taken on a monthly basis at each of the monitoring points specified in the instrument of authorization for a one year period prior to the date of the application.
   3. The following readings taken daily for each directional pattern used during the 30 day period prior to the filing of the application:
      a. Common point current
      b. Base currents and their calculated ratios
      c. Antenna monitor sample current ratios

[5] It is recommended that a field meter of known accuracy with a manufacturer supplied calibrated antenna be used. The antenna dipole length is to be set to distance indicated by the instrument manufacturer's instructions. Data acquisition can be made using a chart recorder or digital storage device.

[6] Typically, pattern modification can occur when the tower face is greater than three feet and/or the antenna operation is in the presence of other antennas or guy wire.

FIGURE 5



TOP VIEW

CONFIGURATION TO PERMIT
FOUR MEASUREMENTS FROM A STATIONARY POSITION

COHEN, DIPPELL AND EVERIST, P.C. CONSULTING ENGINEERS

4.15-7. Configuration to permit four measurements from a stationary position.

FIGURE 6

FIELD STRENGTH MEASUREMENTS

FOR________________________

____

DATE: __/__/__

N __ ° E

FIGURE NO: ________

PAGE NO: ________

ENGINEER: ________

METER: ________

POINT				FM HORIZONTAL				FM VERTICAL		RATIO	POINT DESCRIPTION/GEO COORDINATES
NO.	DIST.	DATE	TIME	dBu	mV/m	DATE	TIME	dBu	mV/m	V/H	

4.15-8. Sample field strength measurement form.

d. Antenna monitor phase readings

e. Final amplifier DC input voltage and current

4. The results of either a partial proof-of-performance (Section 73.154) or a full proof (Section 73.186) conducted no longer than 3 months prior to the filing of the application and the common point impedance at the operating frequency measured at the time of the proof.

Additional sampling system components and configurations found by the Commission to be accurate and stable over a wide range of environmental and operational conditions will be acceptable under Paragraph A above and announced periodically via FCC Public Notice.

Section 5: Video Production Facilities

Introduction by Andy Butler, PBS, Alexandria, VA

The average broadcast station consumes more than 8,000 hours of programming a year. As location is the key to success in real estate, content is the driving force in broadcasting. The ability to create audience pleasing material is in large part what distinguishes broadcasters from pure content haulers like telephone and many cable companies. This ability has never been more important. It is the key to maintaining broadcasting's competitive edge in the battle for eyeballs. The importance of quality content will continue to grow as DTV provides even more opportunities for innovation in content packaging and cross media delivery.

Providing the facilities that support competitive creative efforts is a key task for broadcast engineers. Creating the right combination of tools to effectively support all of the production processes requires a great deal of knowledge in a wide variety of fields. These areas range from mundane, but critical support items, like intercom systems to advanced theoretical topics including photometry and video compression.

In the following section a diverse group of industry experts share their expertise. We begin with a comprehensive overview of facility planning by David Carr. Jerry Whitaker continues with a look at the physics and physiology underlying visual imaging. Peter Gloeggler moves us from the theoretical to the practical with a survey of the options for capturing and manipulating video images. Jay Kuca details the challenges in routing and switching the signal after it's captured while Joe Grega surveys options for storing video.

A key component of effective video usage is efficient storage. Advanced compression techniques combined with the latest recording technology and disc based storage systems are offering flexible, cost effective alternatives to traditional tape based alternatives. Peter Symes, Steve Epstein and Jerry Whitaker bring us the latest in technique and technology.

The production process begins with an idea. Turning that idea into compelling visual images often requires intense manipulation of the captured content. Jeff Mazur looks at some of the latest tools for creating special effects.

Infrastructure is not as flashy as the latest production gadget but a weak support system can destroy the best creative effort. John Weigand and Brad Dick tackle the gritty details of communications, automation and control, while Robert Beach scans the horizon for the latest update on weather radar for TV.

Part of the magic of broadcasting is its universal appeal. Making that magic easily accessible to everyone is a key goal for the industry. Amnon Saloman takes a comprehensive look at closed captioning and other enhanced services.

Creating and capturing compelling images is an art. The medium that stores those images can have a profound impact on their effectiveness. Film has played a major role in television for many years and is still the preferred palette for many video artists. It is also the most universal medium for international program exchange. With all of its strengths, using film effectively on television is a challenge. Richard Bauer looks at the legacy and the limitations.

Program production is expensive. To remain competitive broadcasters constantly seek ways to do more with less. Camera robotics has become an important tool in the quest for efficiency but they also offer quality benefits and expand creative possibilities. Robert Murch explores the efficiencies and the creative bonuses.

Images are reflected light. Lighting paints the canvas of the TV image. The right lighting reinforces and strengthens the message while poor lighting distracts the eye and garbles the communication. Bill Marshall goes beyond the glare for a thorough look at the art of illuminating the image.

The best video production is practiced inside studio walls but the events that are the most compelling content rarely occur in a production studio. From the earliest days, broadcasters have been forced to go out in the field for the best programming. Ned Soseman and Carl Bentz are veterans of the remote production battlefield. They have some invaluable suggestions to make sure you get the content you want without sacrificing your health or your sanity.

Television is entering a new age. One of my favorite colleagues has an insatiable appetite for knowledge and a stubborn determination to understand technology at its very basic levels. After spending a couple of years reading, asking questions and probing he summed things up by noting that "DTV is Different . . . its REALLY Different." David Leathers appreciates that difference and offers us a look ahead with some production considerations for DTV.

5.1
PLANNING A VIDEO PRODUCTION CENTER

DAVID CARR
KHOU-TV HOUSTON, TX
AND MALCOLM M. BURLESON

INTRODUCTION

This chapter contains information that may be used as a guide for general planning and cost estimating of facilities. It is beyond the scope of this chapter to describe all of the many details that go into the design of a facility. Rather, it is intended here to provide hints and reminders of some of the more important considerations that should be addressed.

Before planning the budget for such a project, goals must be set. Why is this project being undertaken? Do we want to provide additional space and if so in what areas? Are we planning on taking on production projects? What kind and what magnitude of production sessions are we envisioning? Do we want to convert our studio operations to robotics? An overall agreement on goals must be reached at before any planning and budgeting can be done.

There are many factors that go into determining the initial planning budget, as well as accounting for the final cost figures when the project is completed. The cost per square foot varies greatly depending on what area is being constructed. The cost of studio or technical space is much greater than that of office space. Advance planning combined with close project coordination and continuous oversight by the owner's representative will result in a well-built, cost-effective facility, constructed on time.

Give the construction or refurbishment of the facility the same study that would be given to a major equipment purchase. Research the facility needs, available architects and contractors. Review local codes, construction specification criteria and cost accounting methods. Hire additional personnel to cover for staff that will be occupied overseeing design and construction.

Construction of broadcast and production facilities is not the same as simple office construction. They require special attention to elements such as control rooms, special acoustic and air conditioning treatment and special technical services not normally found in routine construction. Special consideration must also be given to the backup power generators and other redundant systems to be installed. Years of experience have taught the lesson that numerous large and small broadcast facility projects all have one major thing in common: the need to plan ahead and to watch every detail, or else the barrage of changes will undermine the budget and completion date.

GENERAL PLANNING CONSIDERATIONS

The early planning of a broadcast facility involves factors such as: consideration of the market to be served, site selection, studio requirements, nature of programming, expected hours of operation and available capital.

First and foremost decide whether the studio and transmitter are to be combined under one roof or established in separate locations. Wherever practical, it is more economical to combine the studio and transmitter. While the initial equipment requirements may be less, more importantly, the day-to-day operating expenses are substantially lower. With the plant all under one roof significant savings in heating, air conditioning, building maintenance, travel time and personnel can be realized. When a combined operation is not practical, the second best approach is to place the transmitter and other facilities in a separate building and operate the equipment by remote control from the studio. With a good remote control system, a transmitter site can be selected that provides the best coverage and the studio can then be placed in the most convenient location for business and production. The building requirements at the transmitter can be kept to a minimum, with space only for the transmitting and support equipment, a small work area and the heating and air conditioning system.

Building Codes

Virtually any construction or alteration of a building is likely to be covered by one or more codes. In addition to general building codes, there are electrical, mechanical, plumbing and other specialized codes. The building at a separate transmitter site is likely to be subject to fewer building code restrictions, because it is away from heavily populated areas and fewer personnel are involved in its operation. While various national associations have proposed model codes, individual state and local jurisdictions adopt whatever codes, with variations, they find appropriate. Reputable architects, contractors and subcontractors are familiar with local codes affecting their area as they apply to a specific construction job.

Zoning Regulations

Zoning is another consideration that is dependent on the local jurisdiction and can have a significant impact on selecting a studio location, appearance of construction, access to the facility and placement and size of towers and antennas (microwave and satellite). While it is sometimes possible to have zoning requirements changed to suit a particular construction requirement, this should not be counted upon when considering a new site. However, if a site is particularly desirable, an advance team can be employed to:

- Investigate the zoning regulations
- Determine whether they are enforced
- Consider the attitude of potential neighbors to having the proposed facility constructed nearby
- Propose solutions to the problems the zoning regulations are likely to cause.

If the planned construction is for an expansion of an existing facility, a careful review of local zoning regulations is essential. It may be necessary to increase parking if there is a staff increase. Construction that was grandfathered in the past may be required to meet new requirements if a modification is made.

PLANNING FOR NEW AND RENOVATED FACILITIES

The first step in planning a new facility or renovating an existing one is to develop an outline of requirements that will form the basis upon which future decisions are made. Such an outline for broadcast facility planning might be developed as follows:

Site Selection

- Select a site that provides adequate space, not only for immediate building needs but any expansion anticipated
- Provide adequate space, open areas and parking for employees, guests, and studio audience (if under consideration), and a reasonable surrounding landscape environment, if possible
- Access for delivery trucks and a generous loading area with a loading dock is essential if shipments over 100 lbs are expected on a regular basis. The additional cost for deliveries and the occasional rental of a forklift will be paid for in the first year of dock use
- Accessibility to major (and preferably less congested) roads for mobile units, audience and the general convenience of employees will save time and reduce frustration
- Space for a tower for line-of-sight and electronic news gathering (ENG) microwave and communications antennas will be needed either on the site or on the building itself
- Zoning regulations will affect use of towers, antennas, appearance of construction and facility identification or advertising signs
- Consider a space for a helicopter landing pad on the roof or adjacent to the building for possible use for news and traffic reporting
- Consider various environmental elements such as noise from nearby airports and highways, weather to high winds or snow accumulations and of course, susceptibility to flooding or high water
- Take into account the special technical requirements that broadcast and production facilities have, such as installation of satellite antennas, access for remote vans, microwave paths, telco connections and alternate or standby power.

Building Planning

It is essential to pay particular attention to efficient personnel traffic flow and the relationship of studios, administrative offices, technical areas, craft shops, storage and building utilities, such as power, heating, air-conditioning.

The design, layout and space arrangements of a new facility should include means for easy coordination between such operational functions as programming, operations, and engineering; between on-air talent, dressing rooms and studios; and between newsrooms and their mobile units. Examples include:

- Entrance arrangements for employees, guests, VIPs and audience
- Relationship of studios to loading areas
- Relationship of parking areas to employee and audience entrances
- Storage spaces—area, rooms, closets, nooks and corners
- Physical security inside to discourage unwanted entry and passage throughout the facility
- Controlled access areas for staff
- Outside security to discourage theft and vandalism to the building and employee and visitors cars and outside technical facilities (satellite antennas).

Budget Development Checklist

The following checklist is designed for estimating the scope and cost of station construction.

Preliminary

- Land
- Land tests
- Site clearing and preparation
- Architect and engineering fees
- Permits (zoning & building)
- Special consulting fees (decoration, landscaping, communications)
- Local zoning and building codes

General Construction

- Architectural
- Electrical
- Mechanical
- Heating and plumbing
- Special in-house cabling (TV, telephone, computer, PA)
- Communications (microwave, two-way radios, satellite)

Interior Finish

- Wall covering (fabric or paint)

- Floor covering
- Special studio treatment

Furniture and Fixtures

- Decorative office furniture
- Standard office furniture
- Office area built-ins
- Working area counters, cabinets and built-ins
- Studio fixtures
- Draperies
- Artwork

Telephone System

- Broadcast line facilities
- TWX and facsimile (FAX)
- System configuration
- Office intercom system
- Number of private lines
- System: owned or leased
- Backup arrangements (very important)

Computer Systems

- Administrative
- Technical
- Production
- Backup arrangements (very important).

Security Arrangements

- Control of public (entrances and exits)
- Control of all points of entry (including employees)
- Closed circuit TV systems, security tours
- Controlled access doors and gates
- Separation of 9:00–5:00 areas from 24-hr. areas
- Night lighting and emergency power
- Surveillance of ENG and satellite news gathering (SNG) units and parking areas

Miscellaneous Items (Often Overlooked)

- Copier or printer outlet (may require special power hookup and ventilation)
- Special ventilation for odor producing areas (artwork, craft shops)
- Building maintenance equipment closets
- Drinking fountains and vending machine areas (power hookups)
- Space for air-conditioning subdistribution boxes, fans
- Special waste water treatment for film darkroom processing (local codes)
- Cable access to roof for various communications systems
- Provisions for microwave and satellite antenna mounts
- Power, communications, signal lines to SNG, ENG unit areas
- Use of walls for built-in filing cabinets

Building Construction: General

Here are some of the major elements regarding building construction that should be considered:

- **Official Contact.** Restrict contact with the contractor to only those staff or employees given the authority to speak for the company and authorize changes. Route all official communications with the contractor through one individual, if possible
- **Change Orders.** Any changes in the design must be kept to a minimum if the budget is to be maintained. Changes in design are expensive and time consuming. Virtually any suggestion offered to the contractor by the facility staff will be considered a change by the contractor if change authority is not specifically designated. The extra time spent in careful planning in advance of construction will more than pay for itself with fewer change orders. Coordinate any changes with all departments for a possible overlooked error. Make sure the contractor obtains written approval on all change orders and that the station understands the amount of additional cost involved. Several dozen relatively small change orders can result in many thousands of dollars of budget overruns with very little to show for the changes
- **Document the Construction.** Take pictures of all new construction areas at various stages in construction. Alert the architect that as-built drawings are to be provided. In some cases there is added cost, but advance notice will keep the cost down and the accuracy of the drawings will make future changes easier and less costly
- **Structural.** The selection of the main structural system will take into account local conditions, codes, availability of materials, flexibility for future changes, special loading conditions and spans, degree of fire resistance and effect on fire insurance rates. A careful study of local fire code requirements will result in designing a facility that meets the codes and avoids costly retrofit later (doors, fire control systems, ventilation)
- **Partitions.** Plan the layout of the interior of the building to provide sound isolation in the studio and control room areas, yet enough flexibility for ease of removal or relocation or some of the walls and partitions in office and general purpose areas. In general, the walls needed for good sound isolation will be more permanent than those for offices and other areas. It is generally better to make rooms larger in the initial planning and construction stages and subdivide them later (after construction is completed) if necessary
- **Ceilings.** Ceiling spaces contain virtually all utility services: sprinkler system, heating ducts, water, sewer, power for lights (and more recently outlet power), communications (computer, signaling, security, production video and audio), conduits and wire trays. A high degree of accessibility is required to the ceiling spaces as well as good acoustical quality and ease of cleaning. Attention paid to the ceilings during the design phase will result in easier and less costly access to the space later
- **Wall Materials.** Greater durability is needed in technical and production areas, due to the mounting of various equipment to the walls as well as frequent

moving of equipment and supplies. Install wall guards or rails along all corridors and rooms in which equipment will be transported
- **Floors and Floor Materials.** Most office buildings have concrete floors. Avoid placing any conduit in concrete floors in order to reduce installation costs and increase flexibility later. Different floor types and coverings are needed in different areas. Get the best available that is durable yet will reduce traffic sounds and vibration. It will be costly and disruptive to replace it in the future. Use carpeting in areas where footfalls and dolly wheels are likely to transmit sound to a nearby studio, control room or edit booth. When using carpet remember that it tends to build up a static charge that can damage equipment. There are some grades of carpet that will build up less of a charge under certain temperature and relative humidity ranges. Consider the pros and cons of using raised computer floors in control rooms, videotape and other electronic areas. Computer floors are expensive, require constant maintenance and are difficult and dangerous to work with when installing cables. However, they are an ideal duct for ventilating cool air into the bottoms of racks and consoles. They may be more convenient to use in an area devoted exclusively to computer facilities and may be selectively employed in technical areas. Computer floors look better than overhead cable trays but trays are easier to wire and rewire
- **Studio Floors.** Floors for studios have altogether different requirements. Concrete slab floors are typically covered with either a rubberized surface (higher maintenance) or close fitting tile (lower maintenance, but subject to potholes). Slab floors for TV studios must be extremely level if smooth camera movement is expected. If robotics operations are being considered, a self leveling epoxy floor may be installed. It seems to outlast the other floor surfaces under the rigorous operations of robotics. Robotics tend to be harder on the floor because the pedestals go to the same positions time and again.

Studio floors are often isolated from the building structural members in order to reduce sound transfer. Consider this approach, which is expensive, only if the studio will be relatively close to interior (air conditioning, craft shops) or exterior (aircraft or highway) sound sources, or if a higher degree of sound isolation is required. Floors are more difficult than walls or ceilings to acoustically isolate from building vibration and noise transfer.

Footfalls in corridors will be heard in adjacent studios and control rooms if the studio floor is not isolated or the corridor floor is not carpeted or acoustically treated.

- **Studio Ceilings.** If the ceiling of the studio is directly under a roof, consider what acoustic and thermal isolation may be needed if the site is near an airport, subject to heavy rains, snow or hail or hot sunny days. If a helipad is on the roof low frequency noise isolation will be required
- **Tape Vaults.** Give special consideration to this area that contains the station's major programming resources. Use fireproof construction and limit ventilation to no more than that necessary for proper temperature and humidity control. Install instruments to measure the temperature and humidity. Control access to the vault and keep logs of tape movement
- **Sound Locks.** Sound locks are small vestibules to provide sound isolation between noisy places (such as corridors, newsrooms, offices and control rooms) and studios. The sound lock, when properly used, effectively allows entry or exit from a studio during a broadcast or recording session. Sound locks can have several doors and serve more than one control room or studio. The space used by a sound lock is not entirely wasted. The space above the ceiling can be used for the ballasts for fluorescent lights used in the studio or control rooms
- **Sound Lock Doors.** These special doors must be carefully designed and properly installed. In addition to their superior acoustic performance, sound lock doors must have:

  A. A double glass window with rubber mounting to allow a view of who might be using the sound lock to avoid opening both doors at the same time.
  B. All sound lock doors must open outward from the sound lock so that when a slight vacuum is created it more firmly seals the other doors
  C. The outer door of the sound lock must have a latch and optional lock
  D. The inner studio door must have no latch, that would cause an audible click in the studio, but handles must be provided.

- **Windows.** The main product of a broadcast or recording facility is the sound material generated in its studios. Windows can be strategically placed in studios and control rooms to allow clients to view the activities, for visual communications between the areas, for tours to see the operation without disruption and for employees to see who is working in the studios without having to enter or use the sound lock. The additional expense for properly designed and installed windows will be well worth the effort in the long run. Adding windows after facility construction or renovation is completed is even more expensive because the operation will be disrupted during construction, the studio or control room may have to be rearranged to accommodate the new window and conduits and support members in the wall of interest may have to be moved, modified or removed altogether.

If after construction the window is not needed, it is a simple matter to cover it or add drapes to limit the times when it is appropriate for use.

Also consider that an external window will allow light into a television studio making it more difficult to maintain the desired color temperature due to the sun's constantly changing intensity.

Engineering: General

- **Main Power Service.** A new physical facility will require a new main power service and metering to be constructed by the power company, that may pass the cost on to the customer. An addition to an existing facility may require some construction for the additional power capacity. Research the costs with the power company well in advance of construction in order to reach a cost compromise. After all, studios are good electrical customers. Consider an alternate service entrance to reduce the construction on the main service and provide a partial backup in event of a main service failure. Examine the path the main service takes from the substation. Consult neighbors for indications and frequency of power problems. It is not usual for a second service to be installed if convenient. That is, if one is within a block or so of the studio. However, it may be expensive to have it brought in when it is to be rarely used. It may be better to spend the money upgrading or hardening the main service
- **Power Substations.** Consider separate unit substations for air conditioning and electric heating and other major motor loads (such as elevators), general illumination, studio production lighting and technical facilities. However, arrange for a common metering system to keep the per unit usage costs low
- **Lighting in General.** Use fluorescent lights providing about 100 ft candles in all personnel working areas. Use dimmer controlled incandescent lights in control rooms and other technical areas during production activity and to provide ambiance. Include fluorescent lights in the same areas for installation, maintenance, and cleaning
- **Air Conditioning.** Studios require high volume, low velocity air ducts for sound control. They also require a backup system. If two air handlers are used per studio, one as a main and one as a backup, they will take up valuable space and be quite costly. If you have two studios and are considering a tridundant system, that is three air handlers to serve two studios, the third air handler to backup either of the other two, you will have to provide air ducts that in essence connect both studio together. While this is more economical than the previous option, the acoustics of both studios will be compromised. A third option however is available. Have a specially designed air handler built for each studio. Each of these air handlers should have two coils and two drive motors within the housing. Only one coil and motor is used at a time, but in case of failure a simple switch will change over to the other coil and motor. This is much less expensive and requires less floor space than the first option and will not compromise the acoustics of the studios. Consider a separate system or standby component or cross-connecting arrangement for control rooms and editing suites. Install a separate air purging exhaust system for TV studios
- **Energy Savings Systems.** Novel energy savings techniques should be considered but only with a thorough understanding of the consequences of a system outage
- **Miscellaneous Systems.** Telephone, public address, security's tour, door security, closed circuit TV and fire alarm.

Electrical

The following sample specifications for the electrical system of a television station are based on the *National Electrical Code*, and are intended to show how such specifications may be tailored. Elements may be adapted to other applications, such as facility upgrading as well as radio station construction.

All state and local electrical, fire and safety codes are to be observed.

Area Classification

All electrical installations, materials, and equipment shall comply with the classification *General Purpose* except for hazardous areas that shall be designed for Class I, Group D, explosion-proof conditions. (There are normally no hazardous areas in a broadcast facility.)

Incoming Power Service and Metering

Incoming power service should be high voltage (4.16 kV or 13.8 kV) due to high load requirements. Standby partial or full incoming service shall be provided (if practical) with automatic transfer when normal service fails. One primary metering point shall be provided to obtain best possible utility rates.

Primary Distribution

Distribution within the complex may be high voltage (4.16 kV or 13.2 kV) from a primary switchgear to unit substations located as close as possible to the center of the loads served.

Separate unit substations shall be provided for different type loads, as follows:

- A secondary 120/208 V system to handle equipment and motor loads and a 277 V system for all fluorescent lighting
- A secondary 120/208 V system to handle receptacle, incandescent lighting and small equipment loads
- A secondary 120/208 V system to handle studio production lighting only
- A secondary 120/208 V system to handle technical TV loads only

Each unit substation shall include components as follows:

- Primary compartment with high voltage (HV) fused load break switch
- Open dry-type transformer with a delta HV primary and 120/208 V or 277/480 V, three-phase, four-wire secondary with best possible sound rating of transformer
- Voltmeters and ammeters provided for each phase
- A main secondary air circuit breaker
- Molded case feeder circuit breakers and spares

Secondary Distribution

Power shall be extended from unit substations with cable and conduit to automatic circuit breaker panels and motor control centers.

Motor control centers shall be Class I, Type B, with combination magnetic, full-voltage starting, circuit-breaker-type motor starters or circuit breakers only for 480 V, three-phase operation. Each starter shall have three thermal overloads. Power panels shall be designed for 480 V, three-phase, three-wire service. Panels shall be of the dead-front type with automatic circuit breakers of ampere rating as required.

Panels for receptacle and incandescent lighting loads shall be designed for 120/208 V, three-phase, four-wire service. Panels shall be of the dead-front type with automatic circuit breakers of ampere rating as required. Ground fault detector breakers shall be installed as required by codes.

Lighting panels shall be designed for 277/480 V, three-phase, four-wire service. Panels shall be of the dead-front type with 20 A automatic branch circuit breakers. Panels shall be similar to Westinghouse Type NH1B-4.

Conduit

Conduit shall be rigid steel, asphaltum painted when installed in concrete slabs, below grade and outdoors above grade.

Rigid aluminum conduit shall be used for exposed installation in mechanical equipment rooms, damp locations and locations where exposed to mechanical damage.

Rigid aluminum or steel conduit shall be used for all feeder and subfeeder runs. Steel EMT conduit with compression weather tight fittings shall be used for all other branch circuit wiring indoors and above grade.

Wire

HV cable shall be single conductor crosslinked polyethylene insulated and shielded. Building wire shall be Type THW rated at 600 V, 75° C. No. 12 AWG and smaller shall be solid copper. No. 8 AWG and larger shall be stranded. Fixture wire shall be Type AF, 300 V insulation. Minimum wire size shall be No. 12 AWG, except No. 14 AWG for control wires. Maximum wire size shall be 500 MCM.

Grounding

Electrical grounding shall be provided in accordance with the *National Electric Code*. Equipment enclosures, electrical service, transformer neutrals, outdoor lighting standards and cable shielding shall be grounded.

Insulated bushings and double lock nuts shall be provided at all panel boards and pull boxes in feeder runs and pull boxes shall be bonded through with bare copper wire.

A separate technical equipment ground system shall be provided as required. This technical ground system will help to isolate the non-technical electrical noise from the more critical technical system. The key to this technical ground system is to keep all technical racks and equipment from coming into contact with earth ground. The resistance between the earth and technical ground must be continually monitored during installation. The technical ground should be attached to earth ground at one point and only one point.

An uninterruptable power supply (UPS) may be used to power as much of the technical equipment as possible. This will reduce the time necessary to reboot the software based systems in the plant. If you can afford to put all of the technical equipment on a UPS and you have an auxiliary power generator, you will not lose power during the 15–20 seconds that it takes the generator to take over the load.

Switches, Wiring Devices, Wall Plates and Special Enclosures

Single pole switches shall be 20 A, 120/277 V, ac, quiet type. Duplex receptacles shall be 20 A, 125 V, 2 pole plus U-slot ground. Special outlets to be provided as required.

All wall plates for switch, receptacle, telephone and computer outlets shall be 0.06 in. stainless steel.

Telephone System

Two incoming underground services are required, one for technical use and one for business and office use.

The equipment room for the technical service shall be located close to master control and there shall be a cable-tray tie between the telco equipment room and master control.

The business and office system shall be complete, consisting of conduits from equipment room outlying telephone closets and interconnecting panels and then to the various outlets as required. All installations shall be in accordance with the requirements of the local telephone company.

Interconnecting panels shall be steel with plywood backboard with full opening door, latch, cylinder lock and trim. Each telephone closet shall be furnished with plywood backboard for the installation of distribution equipment. Conduits from the equipment rooms to each destination shall be 3/4 in. minimum.

Telephone and computer outlets shall be 4 in. square with bushed aluminum cover plates.

At least two separately fused ac power circuits from the emergency power distribution system shall be provided in each telephone terminal room.

Modern telephone systems are digital or multiplexed systems often requiring no more than two pair between instrument and telephone panel. To save substantial time and money in the future it is recommended that three or four pair be pulled into each office as a home-run to the panel. This will help to serve the needs of future communications services.

Public Address System

A complete public address (PA) system consisting of amplifiers, loudspeakers and microphones; shall be provided. In some cases this function can be accomplished with the telephone system.

Loudspeakers shall be located in corridors and other strategic locations. System shall be zoned as required.

Fire Alarm System

New codes have made fire alarm and control systems extremely complicated. The system must operate reliably. While it is expensive to evacuate a television facility, stations have been known to burn to the ground. False alarms can be kept to a minimum with knowledge of the system and operational procedures. Fire drills must be held annually and may be planned in advance to reduce the disruption to the operation.

The fire alarm system shall be closed circuit zoned, consisting of control cabinet, gongs, manual stations and automatic fire and smoke detectors. Manual stations shall be provided at each stairway on each floor and at all ground level exterior doors.

Automatic thermal or smoke detectors shall be provided in all areas except where sprinkler heads are installed. Each sprinkler alarm valve shall indicate on the fire alarm panel zone annunciator as a separate zone when activated.

A detailed instruction tour of the system shall be provided by the vendor. A fire suppressant system is to be installed in the master control and videotape areas. One practice dump shall be provided under supervision of the station and witnessed by the fire marshall.

Local codes may require that the building be protected by a water sprinkler system. If you are required to use such a system in the technical area, use a dry pipe system. If you are not required by code to use water you might consider using water in the non-technical parts of the building and a gas system in the technical areas. Remember, however that while the gas system produces less damage to the equipment it simply suffocates the fire but does not cool the area as does water. This means that the fire can start again once the gas action ceases.

Provide hand fire extinguishers at exit doors, in corridors and as required in shop and craft areas and as required by local codes.

Cable Trays and Signal Conduits

A system of cable trays and signal conduits originating from master control shall be provided to studio control rooms, studios, microwave rooms, electronic maintenance shop and computer centers.

In addition, a separate cable tray for microwave waveguide shall be provided from the antenna site to areas where the microwave receivers are located (news room or master control).

Special insulation (Teflon) must be used for all signal cables located in an air plenum and not enclosed in conduit. The purpose of this regulation is to keep a fire from traveling from one part of the building to another by way of a burning cable. Because of the substantial extra cost for this cable, serious consideration should be given to running extra cable ducts between technical and production areas, and areas that might be used in the future as well as conduits to most other locations in the facility.

Studio Production Lighting System

Unit substation and dimmer board shall be located as close as possible to studio served.

Unit substation shall include the following:

- Electrically operated main circuit breaker to permit remote control from studio floor
- Transformer with 6-1/2% taps, three above and three below rated primary voltage, to compensate for secondary voltage variations

Other work shall be as follows:

- Wireway with wiring from load side of dimmers to studio floor patch panel
- Studio grid wireways with load wiring to studio patch panel
- Control wiring from studio control console to dimmer board

In sizing a unit, substations serving dimmer boards, a 50% demand factor may be applied to connect dimmer load.

An *on-air* studio warning light system shall be provided as required. Install *on-air* or *in-use* lights beside (not above) doors to specified control rooms, edit booths and studios. Install *on-air* lights above windows to studios. Install *on-air* beacons in noisy scene construction, storage and staging areas.

Security

The following security systems shall be provided:

- Supervision of all exterior doors on ground level, with the control cabinet in the guard's room
- Closed circuit TV cameras at key positions, with monitors in the guard's room
- Manual nonwired security's tour stations located throughout complex
- Electrically operated gates to control automobile traffic
- Controlled access doors and gates to main office and technical facilities.

Emergency Systems

Power for the emergency system shall be provided with a water-cooled diesel generator set with generator output configuration to be 277/480 V, three-phase, four-wire. The installation shall include (but not be limited to) accessories such as automatic transfer switch, output switchboard, battery starting set, oil storage tank, fuel pump, mufflers and vibration isolators.

Generator set shall automatically sense power failure or 80% under-voltage, start engine, attain and maintain speed and transfer designated emergency load. A manual override of start and transfer of load controls shall be provided.

Provide local transformer with primary delta 480 V, three-phase and secondary 120/208 V loads on emergency supply. Install rechargeable battery operated lights in control rooms, studios, corridors, rest rooms and other areas without windows but which are used by the public or extensively used by employees.

Loads on emergency supply shall include auxiliary lights in the generator room, stairway lights, exit signs, selected corridor lights, telephone and PA systems, selected technical lighting and heating, ventilating and air conditioning loads required for transmission of limited live news programs, network and taped programs. Do not connect the rechargeable battery operated emergency lights to the emergency lighting circuit as they will go out when the emergency power comes on.

Broadcasters should plan for more than mere evacuation procedures. Broadcasters should also plan what must be taken with evacuating staff if the alarm is more than just precautionary. That is, determine what steps should be to ensure the continuation of programming. Decide which documents, logs, tapes, data files and other information should be taken out by staff or immediately stored in fire-proof files (for later retrieval) to protect the business and programming operations of the facility.

Lighting—277 V Fluorescent, 120 V Incandescent

Lighting fixtures shall be completely installed with all required outlet boxes and accessories. Lighting levels shall be in accordance with IES recommendations, with minimum 100 ft candles in working areas.

Fluorescent fixtures shall be used for general illumination. Fixtures shall be with 40 W RS lamps and HP factor ballast, with best sound rating. Fluorescent fixture types shall be as follows:

- Recessed with acrylic lens diffuser to be used in areas with hung ceiling
- Surface or pendant mounted with wrap-around acrylic lens diffuser to be used in stairs and other selected areas with exposed ceiling
- Industrial RLM with porcelain reflectors to be used in mechanical equipment rooms, storage rooms, etc.
- Executive offices and conference rooms shall be provided with dimmer-controlled incandescent lighting using recessed fixtures in addition to fluorescent fixtures for maintenance and cleaning.

Selected walls and art work shall be illuminated with recessed ceiling-mounted incandescent wall-washing fixtures.

Make-up room mirrors shall be illuminated with special bracket wall-mounted fluorescent fixtures. Dressing room mirrors shall be illuminated with special wall-mounted strips with incandescent bare lamps. (Consider the use of special color temperature lights for the dressing room and make-up areas so that the colors applied in make-up will appear the same under the studio lights.)

Outside Lighting

Outside lighting shall include illumination of audience concourses, entrances, parking lots, signs, building exteriors, planters and the like, and shall be installed in accordance with IES recommendations.

Miscellaneous

Outlets for wall-mounted clocks operating on 120 V shall be provided in designated areas. A clock system for use with the master clock in master control, studios, and news areas shall be provided as required. Local office intercommunication systems shall be provided as required. An audio-visual system shall be provided for all conference rooms.

Mechanical

Television studios require special consideration in solving the many problems entailed in the mechanical design due to the high-lighting capacity, noise criteria, air distribution and entrances. Radio broadcasting facilities require many of the same considerations, but some of the problems may not be as severe.

Grounding Systems

A technical ground system is one that grounds all technical equipment and contacts earth ground only at one point. To achieve this care must be taken while installing the technical plant. Equipment racks should not come in contact with earth ground. Teflon spacers may be used to provide such an isolation. Care must also be taken to ensure that the racks do not come in contact with earth ground by way of the ceiling grid. Great care must be taken to ensure that no violation takes place during and after construction.

In some instances it may be necessary to install some racks on earth ground. In those cases you will need to isolate all the inputs and outputs to all the equipment mounted in those racks. An example of such equipment is an earth station controller. It is very difficult and expensive to isolate the controller from the motors on the satellite antenna. It is easier to let the controller remain on earth ground and isolate the inputs and outputs to the antenna system.

Air Conditioning Design Criteria—General

The optimum summer and winter design conditions to be maintained by the air conditioning system is 73° F dry bulb and 50% relative humidity for office and general use areas. However, these conditions may need to be varied depending upon the side of building, whether there are windows, geographic location and in craft and technical areas where temperatures may need to be lower and humidity more closely controlled.

Air Conditioning Loads

Studio and production lighting for television constitute the major portion of the heat gain and can exceed 75% of the total cooling load requirement. The unit lighting load requirement in the production area of the studio can equal 50–60 W per square foot of floor area and in many cases this load can occur in any part of the studio since the production area and audience accommodations usually are flexible.

The studio air conditioning system should be separate from the rest of the building so that it can be turned off or down during nonproduction times. Transmission and solar heat gains are minimal since the exterior walls of television studios are well insulated and windowless for lighting and acoustic reasons.

Another contribution to the cooling load results from occupancy heat gain and the fresh air load. The fresh air requirement should be based on either 15 CFM per person, or the equivalent of one air change of fresh air per minute, whichever is greater. Fresh air quantities must conform with all code requirements.

Methods of Air Distribution and Noise Control

Proper air distribution and air movement in broadcasting studios are of critical importance. Systems should be designed so that within a zone of up to 12 ft above the floor, an air movement of 25 ft per minute (fpm) is not exceeded. Air velocities exceeding the 25 fpm cause drafts and movement of performer's hair and clothing, and stage props that are often built of light materials such as thin canvas and plywood.

The air supply should be introduced at a level above the movable lighting grid system to prevent interference with closely spaced lighting system batten strips. Low-level return grilles located at the perimeter of the studio, in principle, would be desirable. However, due to the nature of studio operation, the grille uses valuable wall space and could be blocked off by the cyclorama curtains or by studio props, that would result in an ineffective return air system. It could also be a possible source of noise generation. Locating the return air outlets at a level above the air supply will tend to relieve the neutral zone before it can heat the ceiling and radiate heat downward. Proper location of return air grilles and maintaining low velocities will reduce the problem of air system short circuiting.

Sound power levels (SPL) and noise criteria (NC) ratings for studios are of utmost importance and unless proper consideration is given to this problem, will result in an acoustically unpleasant studio. Noise level should be within a range of NC 20 to NC 25, so as not to interfere with studio performance, particularly during scenes where there is no conversation and no background sound effects. Duct velocities should be designed for approximately 400 fpm within 10 ft of diffuser or register opening, 525 fpm within 10–30 ft from opening, 700 fpm within 30–50 ft of opening, and 800 fpm within 70–90 ft of opening.

All air ducts (supply and return) should be acoustically lined for sound attenuation and the sound power level of all outlets should be carefully checked to ensure that it does not exceed the decibel rating at the end of the duct run, otherwise it will become additive (logarithmic) and negate a portion of the acoustically treated duct.

All exposed pipes should be insulated and all air ducts should be externally insulated to eliminate reflected sound in studios. Where ducts and pipes pass through walls or floor, the openings should be sealed with acoustical sound-deadening material. Ducts and piping should be suspended from vibration isolators. Where ducts and pipes pass through studio walls, flexible pipe and flexible duct connections of appropriate size should be provided to reduce the transfer of acoustic energy into the studio.

Any hole, pipe or conduit that is not properly treated will allow the transmission of sound into the studio and effectively defeat the other expensive sound isolation measures that have been incorporated. Remember this when making modifications to studio facilities after construction is completed.

Mechanical equipment should be located remotely from the studio to eliminate transmission of sound and vibration. All equipment should be properly supported from vibration isolators. Sound traps for sound attenuation and flexible duct connections to prevent transmission of vibration should be provided for all air handling apparatus.

Hire an acoustic consultant to make measurements of the installed studio air-conditioning system in order to insure the system meets design requirements.

Type of System and Control

Each studio should be served by its own air handling apparatus and should consist of supply and return fans, filters, heating and cooling coils, sound traps and a purge exhaust system.

The arrangement and selection of the component parts of the air conditioning system are highly dependent upon the economics, space conditions and geographical location of the project. Special air flow schemes should also be considered. For example, an *economizer* cycle utilizing 100% fresh air during moderate seasons, or a fixed percentage fresh air system can be used. Preheating coils and reheaters may be required depending upon the percentage of fresh air used, the geographical location and the outside humidity. Heating coils can be steam or hot water coils. Cooling coils should be chilled water coils.

In areas where freezing outside temperatures are experienced, special arrangements must be incorporated to prevent possible freeze-up of preheating coils and chilled water coils. Reheaters are often used in individual areas that have less heat buildup than a larger nearby area.

A separate purge exhaust system should be provided to permit the studio to be evacuated during periods when it is not in use. During the purge operation, the system should provide 100% outside air without attempting to maintain studio design conditions. Where an economizer type cycle is provided, purging of the area can be accomplished by resetting controls to 100% outside air.

The control system should be arranged to control studio temperature, and where facilities are provided for audience participation, additional humidity control should be provided.

The installation of a supervisory data center provides operational supervision of the system and normally includes some remote controls for resetting of space temperatures and humidity, starting and stopping of air handling system, read-out of other pertinent air and water temperatures and alarm indication.

A digital control system should be provided that will allow you to completely monitor and control the system. Water temperature, air temperature, valve position, water flow, fan motors, thermostat settings, room temperature, etc., can all be monitored. The range of individual thermostats can be set at the central monitoring station. Filter conditions can also be determined. Alarm monitoring can also help prevent catastrophic events.

Standby Operation

Master control rooms, videotape rooms, editing rooms and the newsroom generally operate on a 24-hour basis, and an air conditioning failure cannot be tolerated for more than a few minutes. A separate air handling system should be employed to serve various combinations or these areas with provisions for standby equipment or cross-connection in the event primary equipment fails. This can be accomplished by interconnecting the ducts and appropriate dampers, with another air handling system serving a noncritical area in the building such as office areas, thereby permitting the technical areas and newsroom to be satisfied during an emergency period. Another desirable feature to be incorporated in the system is provision for handling 100% fresh air in the event refrigeration equipment becomes inoperative.

Multiple refrigeration units and boilers should be provided so that in the event a single unit becomes inoperative, partial operation can maintain conditions in critical areas.

In the event of an electrical power failure, an emergency generator should start automatically to maintain operation of the boilers, heating pump and air handling system serving the master control room and videotape room.

STUDIO ACOUSTICS[1]

When a microphone is opened in a studio, it is expected to pick up only the voice of the on-air talent. But the microphone also picks up building rumble, office noise, air conditioning noise and unwanted reflections from the room walls. The resulting broadcast signal can sound hollow, dull or boomy instead of bright and clean. Poor acoustics can diminish intelligibility and invite listener fatigue. A studio with good acoustic design and treatment is easy to mic and sounds good without the need for extensive audio processing. The sound from a studio with poor acoustics cannot be fixed with electronic processing.

Creating the ideal studio acoustic environment is often an expensive undertaking. It requires the expertise of an acoustical consultant working in cooperation with the architect, both of whom must be familiar with broadcast and recording studio construction. Studio designers often find it necessary to recommend types of construction that are several times more costly to execute than standard office construction. However, the additional expense will result in studios and control rooms that will be well isolated from intrusive noise and vibrations, and will project a clean and crisp on-air signal from live talent, free of unwanted resonances and echoes.

Estimating Costs

When planning new studios, significant cost reductions can be realized by selecting a site which is removed from street, rail and air traffic and manufacturing equipment. Within the building, keep the studios away from noise generating sources such as pumps and air conditioning equipment. Be sure to consider ceiling height, as large HVAC ducts are required to keep air flow noise at a minimum. If insufficient space exists for the installation of the large insulated ducts, expensive alternatives to solve the acoustic problems will be required.

When initially sketching a studio floor plan, remember that studio walls occupy space and should not be drawn as single lines. All partitions have some thickness that will take up floor space. It is not uncommon for some acoustic partitions employing internal air spaces to exceed a 16 in. thickness. If space is not a premium, such constructions are generally more economical than thinner partitions with similar transmission characteristics. Absorptive surface treatments can easily add an additional 4–6 in. to each side of the wall profile.

The control of the transmission and reverberation of sound are two acoustics disciplines which apply to broadcast studio and control room environments. They are independently designed and require completely different methods and materials in creating studios suitable for broadcast and recording. It is important that each be understood, since confusion between the two can aggravate an existing problem.

Controlling Sound Transmission

Sound transmission occurs when a source of sound is carried from one area to another. The vehicle of transmission may be a direct path, a flanking path around an object or a path carrying vibration. The most economical and practical means of limiting the transmission of noise usually requires the control or reduction of it at or near the source. Disturbances with low frequency components are the most difficult and expensive to contain. Completely enclosing the noise source in a highly absorptive room, and isolating its vibrations from the building structure, will limit the ability of noise to intrude upon other areas.

It is often cheaper to isolate the sound at the source than to isolate several rooms from the building structure. For example, if footfalls on a cement floor adjacent to or above the studio are causing problems, try carpeting those areas before attempting to increase the transmission loss of the studio construction. Where treatment at the source is not practical, the most effective way of isolating an area against noise intrusion is to employ *floating construction*. With floating construction as shown in Figure 5.1-1, the noise sensitive

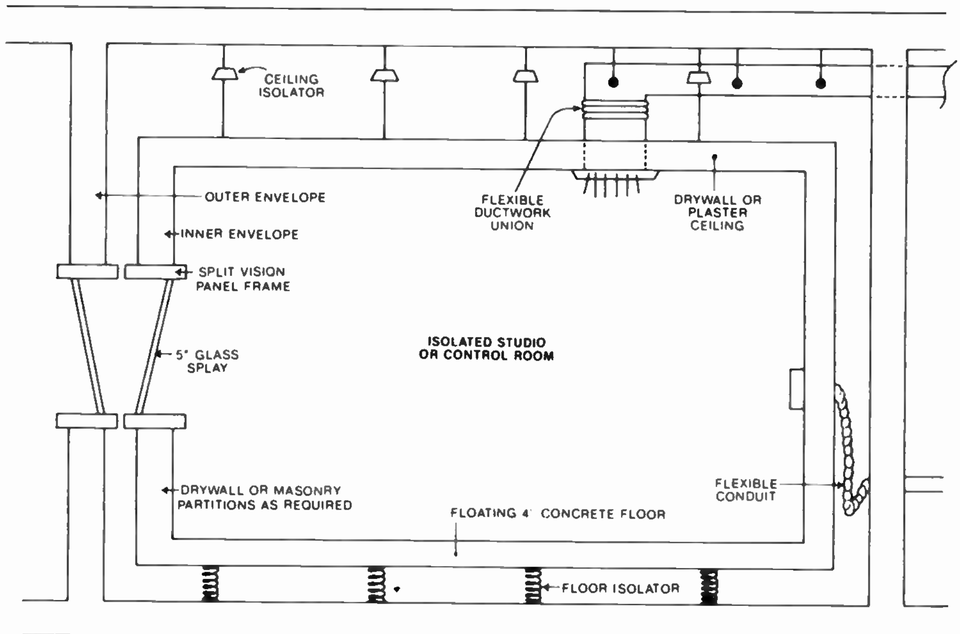


Figure 5.1-1. Floating construction schematic showing mechanical isolation between inner and outer rooms.

area is actually a room built within a room. The important aspect of this type of construction is that mechanical contact between the inner room and outer room is kept to a minimum, with the construction of both rooms creating as complete and uninterrupted an envelope as possible. This is most commonly achieved by resilient floor supports and suspending the ceiling of the inner room on springs or neoprene isolators, hence the term *floating*.

Floating studios in radio stations generally are simpler because the inner and outer walls have fewer connections between them. Television studios, however, have ceilings that are significantly higher and resilient sway brace connections between the inner and outer walls may be required to stiffen the construction and prevent buckling. Television and recording studios often have a variety of utilities provided in the studio, such as electrical lighting and water connections, in addition to the audio circuits that must be isolated. In addition, there are more windows and doors in television studios that must receive special attention.

Depending on the intensity and proximity of the noise source, and the degree of residual noise that can be tolerated, some or all of the aspects of floating construction can be employed in designing a broadcast studio or control room. The resiliently isolated floor and ceiling is necessary to prevent structure-borne sound and vibration from entering the interior space. The double walls with the resulting air cavity between them are actually more effective than a single partition of the same mass. In order to preserve this advantage, no service such as air conditioning ducts, plumbing, or electrical conduit should be allowed to make a solid mechanical contact between the outer and inner rooms. The use of flexible connections within the cavities is extremely important.

When spring or neoprene isolators are employed for vibration isolation, the effectiveness of the isolator is measured by its nominal deflection, the relation of its natural period to that of the disturbing vibration and the stiffness of the structure supporting it. Therefore, isolators with the proper characteristics must be specified. Isolators that are too lightly loaded will provide little isolation, while isolators that are too compliant may cause instability, or bottom out. If the isolation system exhibits a natural resonance equal to the disturbing frequency, the transmission of the disturbing frequency can actually be amplified.

One often overlooked source of vibration that may induce unwanted disturbances into adjacent areas is the control room monitor speaker system. Suspending the loudspeaker enclosure from the wall or ceiling on a resilient mounting will significantly reduce the structure-borne transmission of sound from the monitor.

Acoustic Materials

The materials used for the reduction of sound transmission in walls, ceilings, and floors are typically heavy, dense and thick. Factors to be considered when designing a barrier for maximum transmission loss include the stiffness, resonance, mass, isolation, cost and construction details of a partition. Typical acoustic partitions may be composed of brick or masonry block, poured concrete, lead, multiple layers of gypsum board supported by metal studs or a combination of these materials. Regardless of the materials used, the most efficient barriers employ an integral airspace or cavities filled with a damping material such as fiberglass bats.

Isolation at the lower frequencies is more difficult and expensive and usually requires stiffer, thicker and heavier construction than that needed for higher frequencies.

The single most important aspect in the construction of acoustical barriers is sealing. Regardless of the materials used, the fit of the individual components with each other and the existing structure must be tight. All joints should be filled and sealed with a resilient, nonhardening caulk. The smallest openings can have serious consequences, and render otherwise expensive construction no better than a far cheaper counterpart.

Table 5.1-1 shows how the sound insulation value (R value) of a hypothetical partition at some mid frequency, is reduced due to acoustic leakage through various size openings. Example: Partition with R value = 60 dB

Controlling Reverberation Time

Separate and distinct from sound and vibration isolation, is that part of acoustics which governs how a room sounds when a sound originates in it. This is almost entirely a function of the room size, proportion and the ability of its contents to reflect, diffuse or absorb sound of differing frequencies. To complicate matters, all these parameters interact with each other in determining the reverberation characteristic of a room.

The major measurable acoustic characteristics are a combination of:

- The time that it takes a sound to decay 60 dB within a room, once the source of that sound is terminated (T60)
- How T60 differs with frequency
- How uniform the decay rate is
- The ratio of early and late reflections
- The natural room resonances (modes).

Room tuning is both an art and a science, and research continues to more fully understand the process. While divergent philosophies are often involved, most consultants agree on the following guidelines for broadcast and recording studios:

- The smaller and more symmetrical a room, the more noticeable its undesirable resonances. This is why many television announce booths sound more like stuffy little phone booths
- Avoid exceptionally long and narrow proportions, square rooms, rooms with concave walls and rooms with a ceiling height equal to the height or width. Splayed walls and ceilings are dramatic, but necessary only as opposing surfaces that cannot be covered with mid and high frequency absorptive material, such as large vision panels and glass doors.

Table 5.1-1

Reduction in the sound insulation (R) of a partition due to acoustic leakage through various size openings.

Size of Opening	Resultant R
0.0%	60 dB
0.1%	30 dB
1.0%	20 dB
10.0%	10 dB
50.0%	3 dB

Controlling the T60 of a room yields the most dramatic results. A medium size radio studio with a T60 of approximately 0.3–0.4 seconds from 100 Hz to 6 kHz will yield a pleasant acoustic environment.

Unlike the massive solids that are used for acoustic isolation, the most common absorptive materials are light and porous. The most common materials that are commercially available for absorbing sound are carpeting, acoustic tile ceilings and fiberglass or polyurethane foam wall panels. The difficulty in using these materials in broadcast studios is that they provide only mid and high frequency absorption.

The exclusive and excessive use of these materials can cause a studio to become boomy, by having a long T60 at low frequencies in proportion to the short T60 at the higher frequencies which these materials absorb.

An acoustic consultant can specify the design of resonant slot, hole and panel absorbers, as well as extra thick mineral fiber materials to absorb low frequency sound. Boominess can also be decreased by adding a thick fiberglass blanket above a lay-in tile ceiling and by using commercial absorptive materials in their thickest available form. Applying 3 or 4 ft widths of these materials with 2 or 3 ft spacings between them may also help balance the T60 of low and high frequencies. However, to avoid reflective echoes, no hard untreated surface should ever oppose another, either parallel or at an acute angle to it.

In combination control rooms (*combo* studios), often used in radio stations, mechanical equipment such as cartridge tape machines and reel-to-reel tape decks should be surrounded with as much absorptive material as practical. This will help absorb some of their mechanical sounds that might otherwise be reflected toward the host's microphone. Also, avoid placing the console microphone position too close to a vision panel. Whether omnidirectional or cardioid, any conventional microphone requires a free field behind it. A reflective surface behind the microphone such as a window or script panel will color the sound it picks up from the front.

Large television studios may derive as much of their acoustic characteristic from their sets and backdrops as from any materials purposely installed for acoustic purposes. However, because any portion of a studio wall may be exposed at one time or another, it's a good idea to cover the walls with absorptive mineral bats. The bats should be protected by a wire mesh, to keep them from disintegrating when props and sets are stored up against them.

The ceiling, above the lighting grid, should also be heavily absorptive, especially since large portions of the floor will remain reflective. It is important to keep the T60 of the television studio quite low to minimize the transmission of camera and crew noises, and to permit greater distances between talent and microphone without an off-mike quality. Hard, concave, acoustically reflective sets must be avoided, since any combination of these shapes tends to reflect unwanted sound toward the talent microphones in front of them.

The most reliable rule of thumb in acoustics is that treating low frequency problems is always more difficult and expensive than mid and high frequency work. If the budget is tight, always assign top priority to sound transmission considerations. Once the facility is built, little can be done to make up for economies made in the basic construction, while considerably more flexibility for improving interior room acoustics will remain.

Sound Locks

Once the design of the partitions has been decided, an entrance with the same sound insulation characteristic will be needed. There are few practical doors commercially available which meet or exceed the R values of the most common partitions used in broadcasting. Therefore, a sound lock entrance scheme is often required, that will attenuate sound through two doors and an intervening air space.

The sound lock is a small vestibule between the studio and a hallway or other room that provides access to the studio as shown in Figure 5.1-2. By entirely covering the walls of the sound lock between the studio and external doors with absorptive material, the efficiency of the sound lock will increase, decreasing the effect of opening one door. The use of a sound lock not only reduces the insulation requirement for each door, but also provides passage in and out of the studio or control room without exposing it to the full noise and disturbance of an adjoining area. Doors with good acoustic characteristics are available either in metal or wood, both of which feature a sound retardant core.

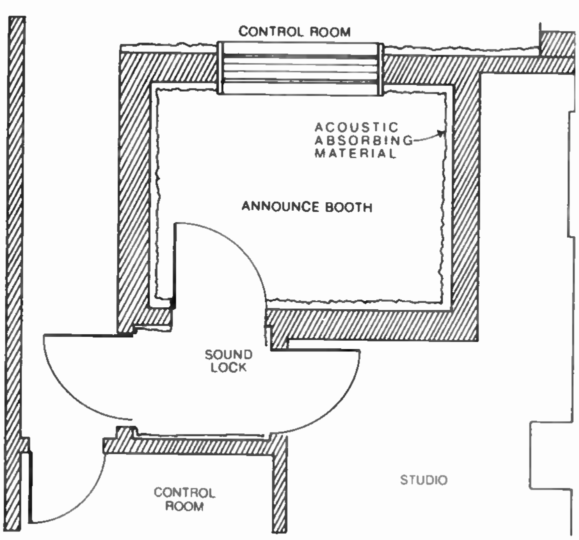


Figure 5.1-2. Soundlock entry to a small announce booth.

Like the partition requirements, the sound lock door must seal tightly within its frame. Lack of a good seal wastes whatever investment is made to procure a good sound rated door. Commercially available compression and/or magnetic type seals are recommended at the head and jambs of the door, while a mechanical drop seal and step saddle should be employed at its threshold. Some expensive prehung doors feature integral seals and can lift hinges that actually lower the entire door into place tightly against the threshold saddle when it is closed. The greatest advantage of these doors is that they never have to be pushed shut to make a good seal all around the door. All other types require an oversize door check (pneumatic or hydraulic closer) and a latch to properly compress the seals when the door is closed.

Windows

Studio windows require special consideration. In order to match the transmission loss characteristic of the partition surrounding it, a studio window must be double glazed. The two panes of glass should be of different thicknesses to minimize the coincidence of each pane's deficiencies. The glass should be mounted in a resilient neoprene or felt lined channel and caulked tight. Some provision should be made for removing either piece of glass for replacement or cleaning. The glass should be mounted with a 5° outward splay at the top to reduce acoustic and visual reflections.

Where a high degree of isolation is required, acoustic laminated glass should be employed. Each pane of this glass consists of alternating layers of plastic and glass and it is lighter, less resonant, and a better acoustical barrier than standard plate glass.

Welded hollow metal frames are preferred over wood for both doors and windows for their stability. However, the void within them should be packed with either a cement-like or mineral fiber filler. It is important that the frames not connect the vertical partition sections in floating room or double wall construction.

Modular Rooms

In large cities where the cost of construction labor is high, small studios, announce booths and control rooms may be economically built from modular components. The modular rooms are constructed from 4 in. thick, prefabricated hollow steel panels. A modular room usually includes a self-supporting ceiling and an integral floor. The entire room is assembled on rails that can be isolated with springs or neoprene from the building structure. Most companies offering modular rooms also include double and triple partitions, doors, windows, conduit, ventilation ducts and a performance guarantee as available options. Because modular rooms can be disassembled and relocated, they are not usually considered leasehold improvements and may offer certain financial benefits in addition to being somewhat portable.

ACOUSTICS SUMMARY

Regardless of the design, the installation contractor must be familiar with the stringent requirements of acoustic construction. The efforts of the general construction, electrical and mechanical trades must be coordinated by the architect both in the design and execution of the studio project to assure that no conflicts compromise the acoustic plan. The job should be inspected during construction by the acoustical consultant, to check for potential leakage paths brought about by existing field conditions, and conformance to acoustical details. If the station or studio is to be situated in an office building, or other shared location, consider that other tenants may be annoyed by high-level monitoring, tape rewind noises and other sounds peculiar to broadcasters. Therefore, during the design phase, if the acoustic consultant recommends protection for adjacent building tenants, the extra effort will result in happier neighbors.

ESTIMATING FACILITY COSTS

Estimating the cost of a broadcast or production facility is a complicated process. In addition to the technical equipment costs, that can be easily estimated, there is the cost of the physical facility construction or renovation that varies according to:

- The area or location in which the facility is to be built or renovated
- The quality of materials and level of attention to detail
- The time allowed for construction and time of year of construction
- The employment of energy saving techniques for long term savings
- Whether an architect, consultant or station staff will be supervising the entire project.

The costs of construction continue to increase. Materials and labor must be examined carefully by competent architects and construction engineers for the specific structure and area involved, if there is to be any accuracy in budgeting the project.

A major factor contributing to cost overrun on projects is inadequate initial plans and the resultant change orders during construction. Therefore, the owner should provide a supervisor to be directly involved in the project on a daily basis. While it is not necessary for the supervisor to be a construction expert, it behooves management to select someone who is skilled in project management, has a general understanding of the project and the time to devote to it.

Standardized Construction Specifications

Most contractors employ standardized methods for estimating costs of construction and breakout costs in a standardized manner. An example of a standardized list of the different construction areas, disciplines and trades is the *Masterformat* published by Construction Specifications Institute. All construction is divided into numbered 16 super-categories and each super-category is subdivided into more specific construction type and third and fourth levels of specifications. For example, a contractor would refer to Section 16780 for Television Systems (Broadcast Video Systems), under the sub-head of 16700 for Communications, under Division 16 which is Electrical.

Another example of the use of the standardized codes would be for the major heading of 15 for Heating, Ventilating and Air Conditioning system under which are sub-headings (or products) including 15550—Heat Generation, 15680—Water Chillers, 15750—Heat Transfer, and 15850—Air Handling.

When working with competitive bidding on the project the standardized code numbers provide the means for comparing the costs for a given kind of work.

CONSTRUCTION EXAMPLES

The following two examples illustrate construction costs for television facilities in large cities. The costs shown should be increased by some amount to reflect the increase in construction costs since these projects took place. In each of the cases, a general contractor was retained and coordination was handled by station staff engineering personnel. Electrical costs include power to all equipment, but not wages paid to station staff technicians who installed the broadcast equipment.

Example A. New Studio Facility

For this addition to the studio facility there were four basic requirements that had to be met (see Figures 5.1-3 and 5.1-4):

- Provide Television Station offices
- Provide expansion for the News Department
- Provide two studios and ancillary facilities capable of producing local newscasts as well as increased local station production
- Provide technical facilities that would meet future needs.

KHOU-TV, the Houston CBS affiliate underwent an expansion and remodeling project from August, 1993 until July, 1995. The 24,000 sq. ft addition and the remodel of the existing 36,000 sq. ft facility took two years to plan and two years to build.

Part of the project included the expansion 912 sq. ft. to one of our studios. An exterior wall had to be removed to allow for the increased studio space. Also included in this portion of the expansion was the acoustical applications as well as the installation of the self-leveling epoxy floor.

The asbestos found in the existing facility was abated as part of this project. The costs listed below are in addition to the cost of equipment. The technical support and equipment installation totaled $1,250,000, while the equipment purchased was $3,817,150. If this amount seems high, keep in mind that this analog facility was converted into the country's first all digital television station.

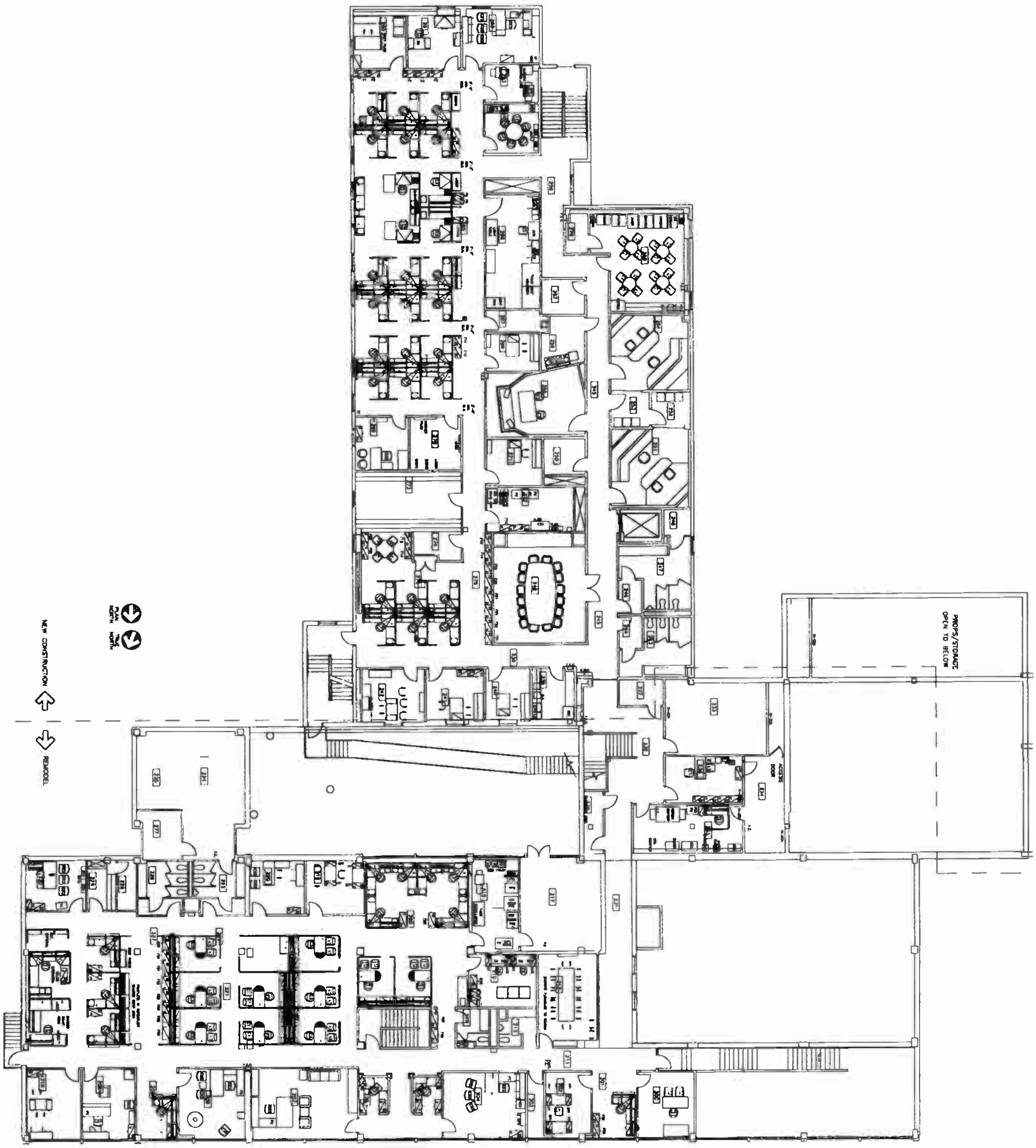


Figure 5.1-3. KHOU studio facility second floor.

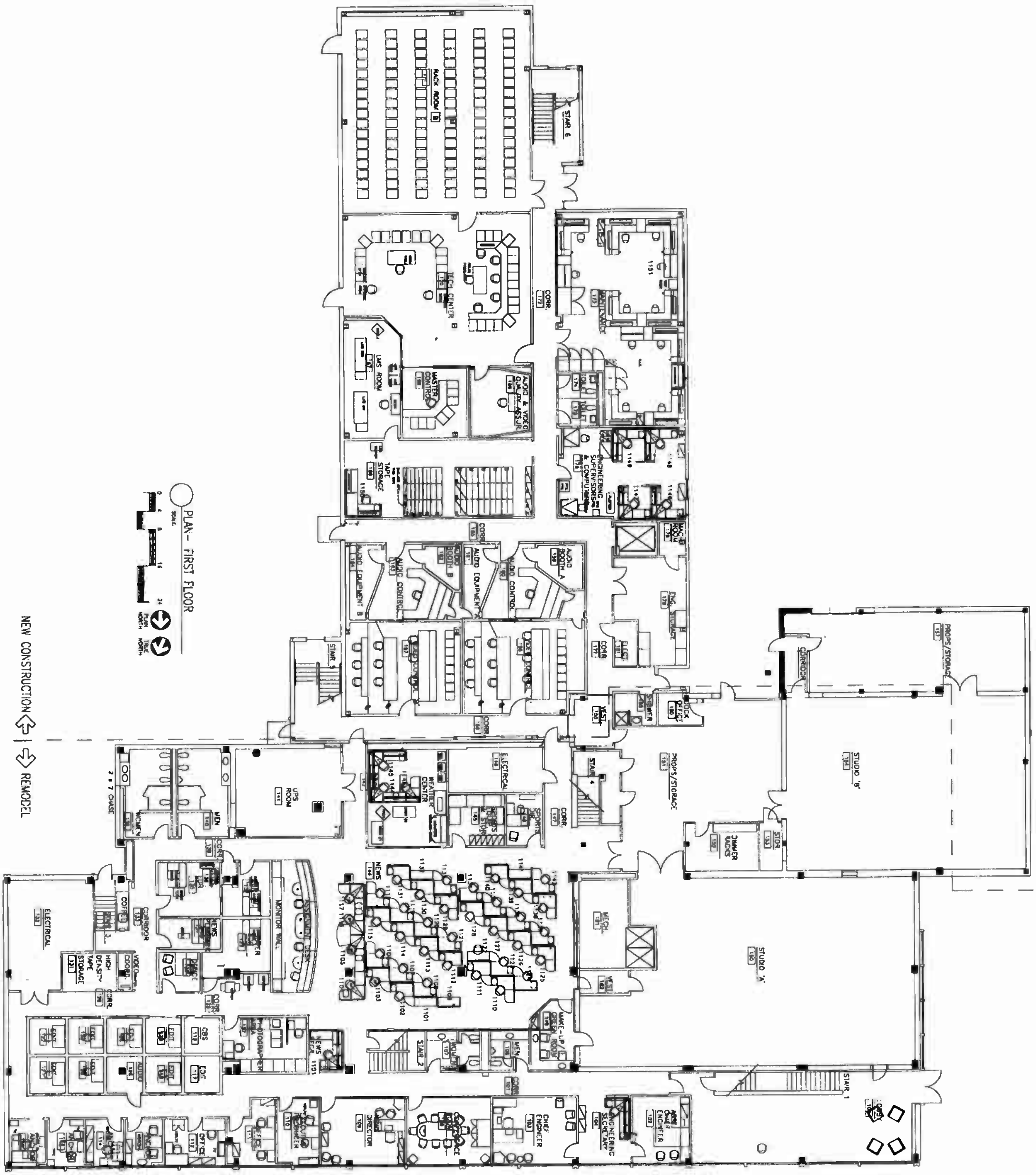


Figure 5.1-4. KHOU studio facility first floor.

Example A. Cost for New Building and Renovation of Existing Building and Expansion of Studio

Activity	Cost	Per Sq. Ft.
Land Acquisition	$ 108,500	
Site Survey & Site Work	$ 115,000	
Architectural Fees	$ 639,700	
Construction of New Space	3,328,600	$139.00
Renovation of Existing Space	2,373,300	66.00
Asbestos Abatement	212,700	
Furniture	601,500	
Contingency	174,500	
Studio Expansion	503,400	552.00
Auxiliary Power Generator	214,300	

Example B. Transmitter Building

In this example a transmitter site is to be relocated from downtown Houston to a site outside of town and a new transmitter building and towers are to be constructed (see Figures 5.1-5 and 5.1-6). The following goals were set:

- Relocate from downtown site which was being surrounded by buildings taller than the existing transmitting antenna
- Locate in same general area as the existing Houston TV stations
- Permit an antenna height of 2,000 ft AMSL
- Provide a Principal City Coverage over the same area as the old transmitter site.

A site was found that met FAA approval for a 2,000 ft AMSL antenna height. Two towers were to be constructed 100 ft apart. The foundation design included both towers and the transmitter building. The transmitter and antenna were selected to provide the best compromise between minimizing primary power consumption and providing 5 MW radiated power.

Example B Cost for New Transmitter Building—2,300 Sq. Ft.

Activity	Cost	Per Sq. Ft.
Structural	$29,000	$12.72
General Construction	$77,000	33.77
Decoration	$ 6,000	2.63
Heating & Ventilating	$15,000	6.59
Electrical (including Auxiliary generator, pad and roof)	$67,898	29.78
Plumbing	$10,000	4.39
Architect, Elect. & Mech. Engineering	$14,730	6.46
Special Woodwork & Built-Ins	$ 4,000	1.75
		$98.09

Typical Construction Cost Ranges

Activity	Cost Per Sq. Ft.
New Studios and Offices (new structure)	$82 to $138
Conversion for Studios & Offices (in existing structure)	$40 to $65
Transmitter Building (new structure)	$61 to $98

When divided into separate elements the breakout is as follows:

Activity	Cost Per Sq. Ft.
Demolition and Site Work	$1.80 to $3.00
Structural	$9 to $14
General Construction	$18 to $50
Heating, Ventilating and Air Conditioning	$9 to $16
Electrical	$5.40 to $18.00
Plumbing	$1.35 to $4.00
Architect and Engineering Fees	$2.25 to $7.00
Special Acoustical	$2.70 to $5.00
Woodwork and Built-Ins	$1.08 to $3.00
Office Furniture	$2.70 to $7.00
Decoration	$3.60 to $7.00

Electronic equipment installation and wiring is not covered by any of the previous estimates. A good

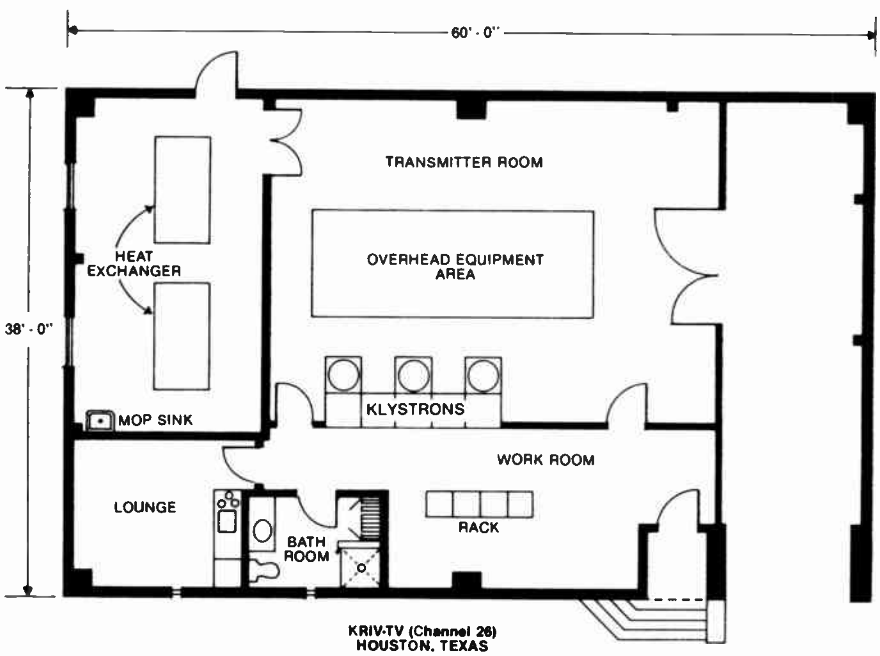


Figure 5.1-5. Transmitter building.

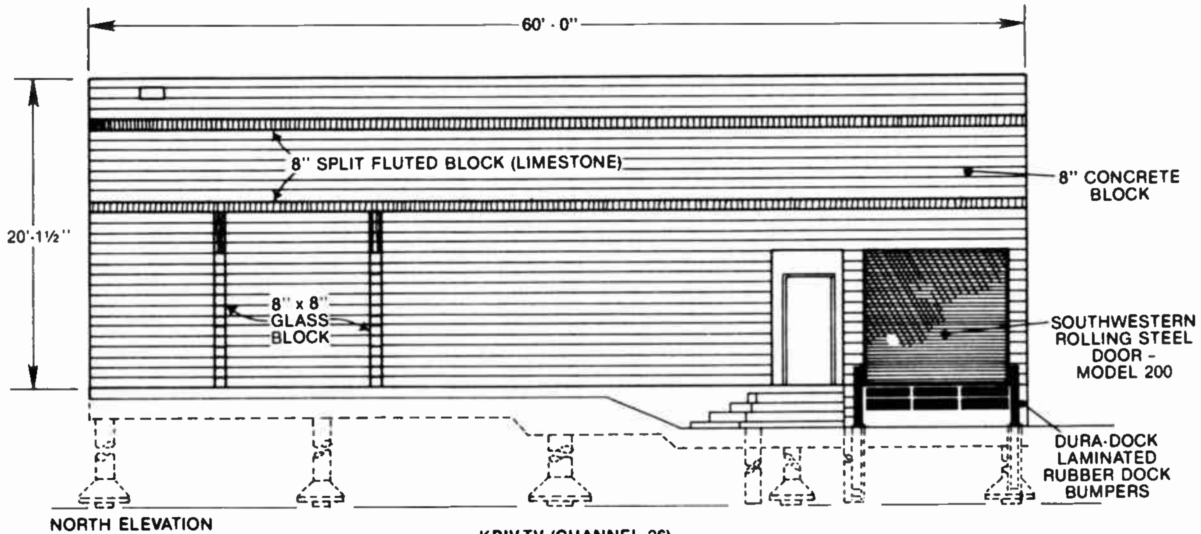


Figure 5.1-6. Transmitter building.

general rule is to allow 15–25% for this in addition to basic equipment cost. Variations are due to location, personnel and working rules.

CONCLUSION

The construction or renovation of a broadcast or recording facility is a major undertaking. Only the briefest explanations of the various elements and concerns were provided in this chapter. It behooves station management to become intimately involved in the construction details or to have a knowledgeable employee or consultant supervise the work. Advance planning and coordination with all departments and personnel will help design the best facility and smooth the construction and the transition to the new facility.

REFERENCES

1. This section was contributed by Alfred W. D'Alessio, Northeastern Communication Concepts.
2. Example A was contributed by David Carr, KHOU-TV, A. H. Belo. Example B was contributed by Richard J. Anderson, Metromedia Television.

5.2
PRINCIPLES OF LIGHT, VISION, AND PHOTOMETRY

BY JERRY WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

Vision results from stimulation of the eye by light and consequent interaction through connecting nerves with the brain.[1] In physical terms, light constitutes a small section in the range of electromagnetic radiation, extending in wavelength from about 400 to 700 nanometers (nm) or billionths (10^{-9}) of a m. (See Figure 5.2-1.)

Under ideal conditions, the human visual system can detect:

- Wavelength differences of 1 millimicron (10 Å, 1 Angstrom unit = 10^{-10} cm)
- Intensity differences of as little as 1%
- Forms subtending an angle at the eye of 1 arc minute, and often smaller objects

Although the range of human vision is small compared with the total energy spectrum, human discrimination—the ability to detect differences in intensity or quality—is excellent.

Sources of Illumination

Light reaching an observer usually has been reflected from some object. The original source of such energy typically is radiation from molecules or atoms resulting from internal (atomic) changes. The exact type of emission is determined by:

- The manner in which the atoms or molecules are supplied with energy to replace that which they radiate
- The physical state of the substance, whether solid, liquid, or gaseous

The most common source of radiant energy is the thermal excitation of atoms in the solid or gaseous state.

The Spectrum

When a beam of light traveling in air falls upon a glass surface at an angle, it is *refracted* or bent. The amount of refraction depends upon the wavelength, its variation with wavelength being known as *dispersion.* Similarly when the beam, traveling in glass, emerges into air, it is refracted (with dispersion). A glass prism provides a refracting system of this type. Because different wavelengths are refracted by different amounts, an incident white beam is split up into several beams corresponding to the many wavelengths contained in the composite white beam—thus is obtained the spectrum.

If a spectrum is allowed to fall upon a narrow slit arranged parallel to the edge of the prism, a narrow band of wavelengths passes through the slit. Obviously, the narrower the slit, the narrower the band of wavelengths or the sharper the spectral line. Also, more dispersion in the prism will cause a wider spectrum to be produced, and a narrower spectral line will be obtained for a given slit width.

It should be noted that purples are not included in the list of spectral colors. The purples belong to a special class of colors; they can be produced by mixing the light from two spectral lines, one in the red end of the spectrum, the other in the blue end. Purple (magenta is a more scientific name) is therefore referred to as a *nonspectral color.*

A plot of the power distribution of a source of light is indicative of the watts radiated at each wavelength per nanometer of wavelength. It is usual to refer to such a graph as an *energy distribution curve.*

Individual narrow bands of wavelengths of light are seen as strongly colored elements. Increasingly broader bandwidths retain the appearance of color, but with decreasing purity, as if white light had been added to them. A very broad band extending generally throughout the visible spectrum is perceived as white light. Many white light sources are of this type, such as the familiar tungsten-filament electric light bulb (see Figure 5.2-2). Daylight also has a broad band of radiation, as illustrated in Figure 5.2-3. The energy distributions shown in Figures 5.2-2 and 5.2-3 are quite different and, if the corresponding sets of radiation were seen side by side, would be different in appearance. Either one, particularly if seen alone, however, would represent a very acceptable white. A sensation of white light can also be induced by light sources that do not have a uniform energy distribution. Among these is fluorescent lighting, which exhibits sharp peaks of energy through the visible spectrum. Similarly, the light from a monochrome (black-and-white) video cathode ray tube is not uniform within the visible spectrum, generally exhibiting peaks in the yellow and blue regions of the spectrum; yet it appears as an acceptable white (see Figure 5.2-4).

[1] Portions of this chapter were adapted from: Whitaker, Jerry C., *Electronic Displays: Principles, Design and Applications*, McGraw-Hill, New York, 1993. Used with permission.

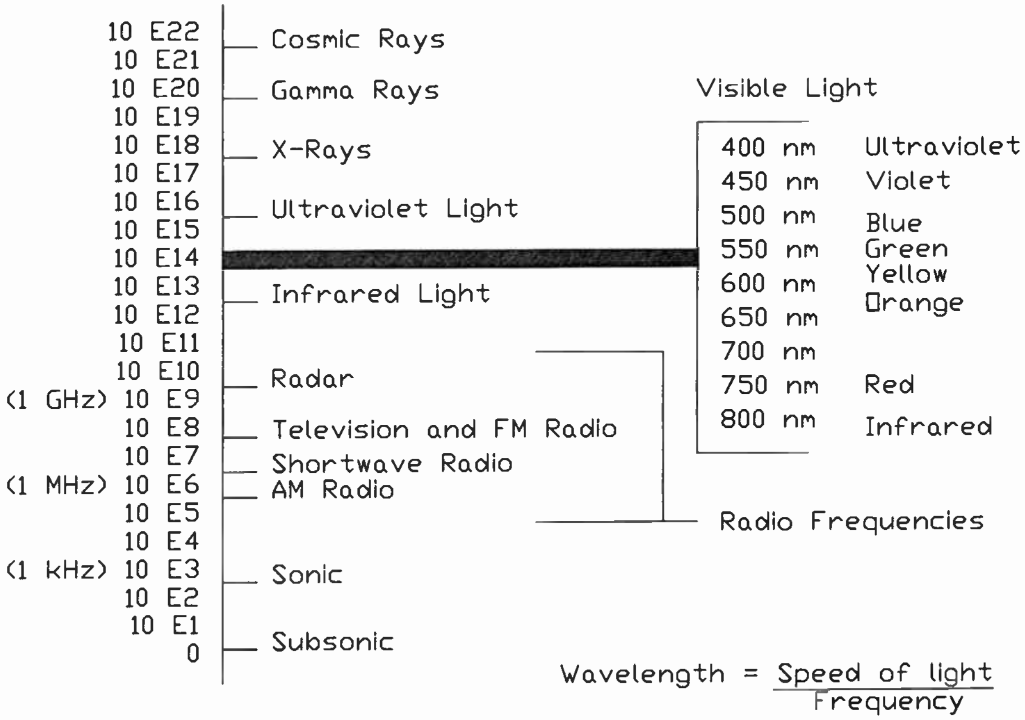


Figure 5.2-1. The electromagnetic spectrum.

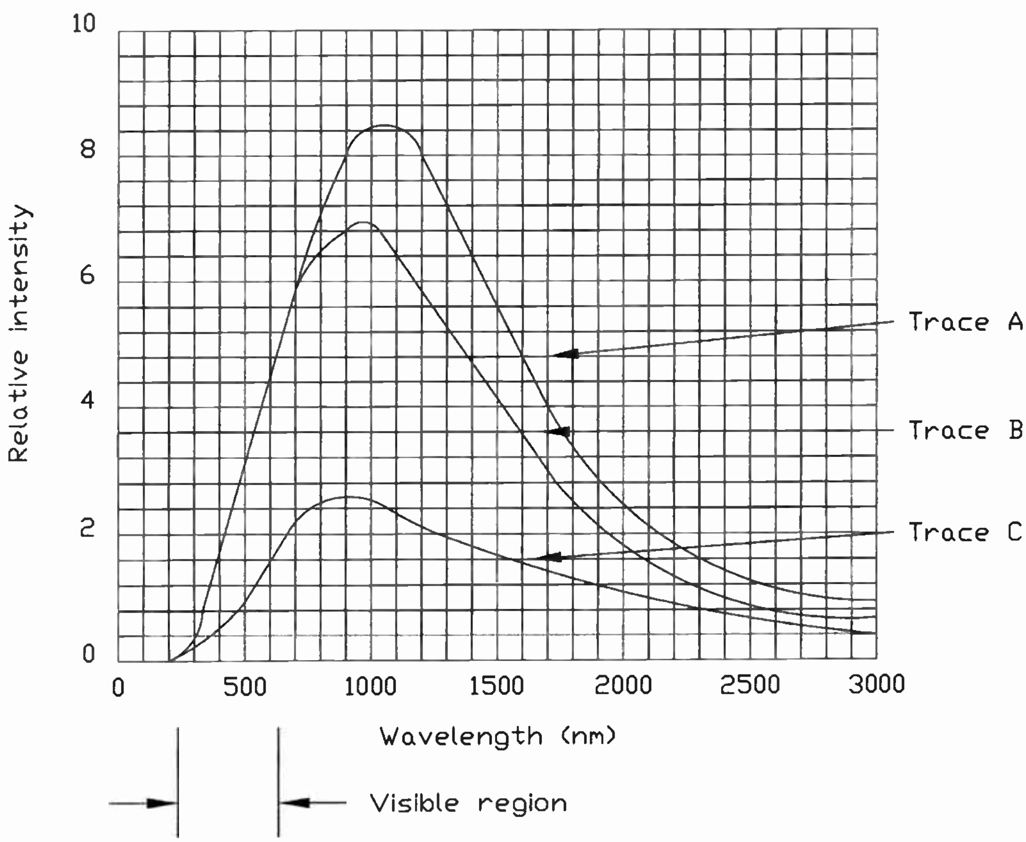


Figure 5.2-2. The radiating characteristics of tungsten: (trace *A*) radiant flux from 1 cm^2 of a blackbody at 3000°K, (trace *B*) radiant flux from 1 cm^2 of tungsten at 3000°K, (trace *C*) radiant flux from 2.27 cm^2 of tungsten at 3000°K (equal to curve *A* in the visible region).[6]

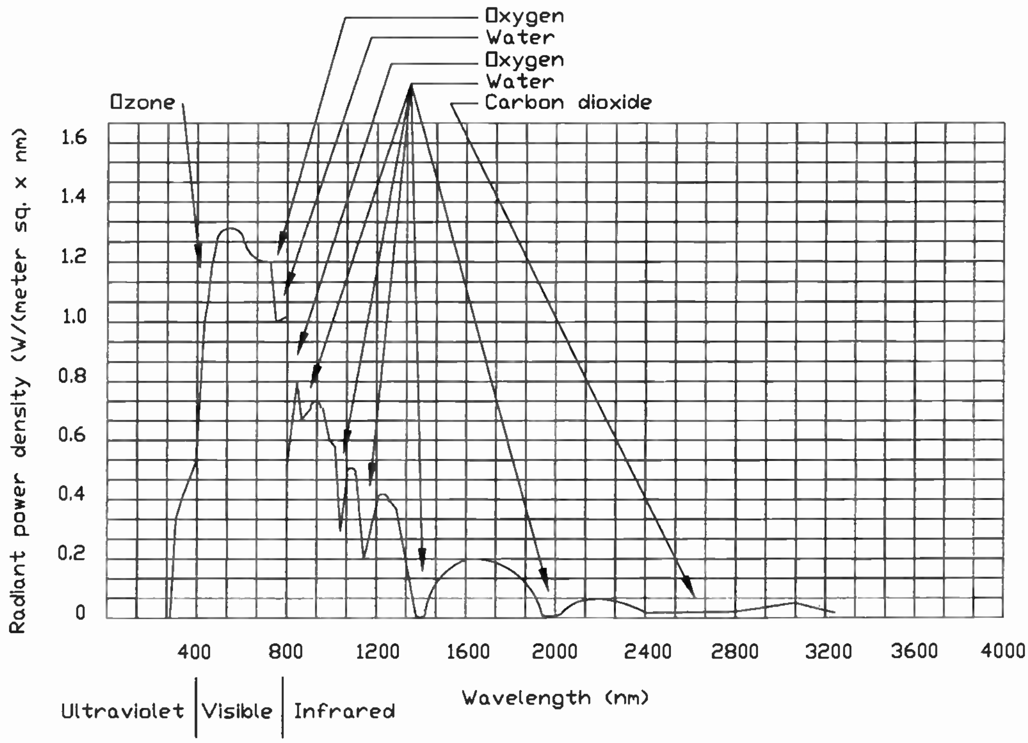


Figure 5.2-3. Spectral distribution of solar radiant power density at sea level, showing the ozone, oxygen, and carbon dioxide absorption bands.[6]

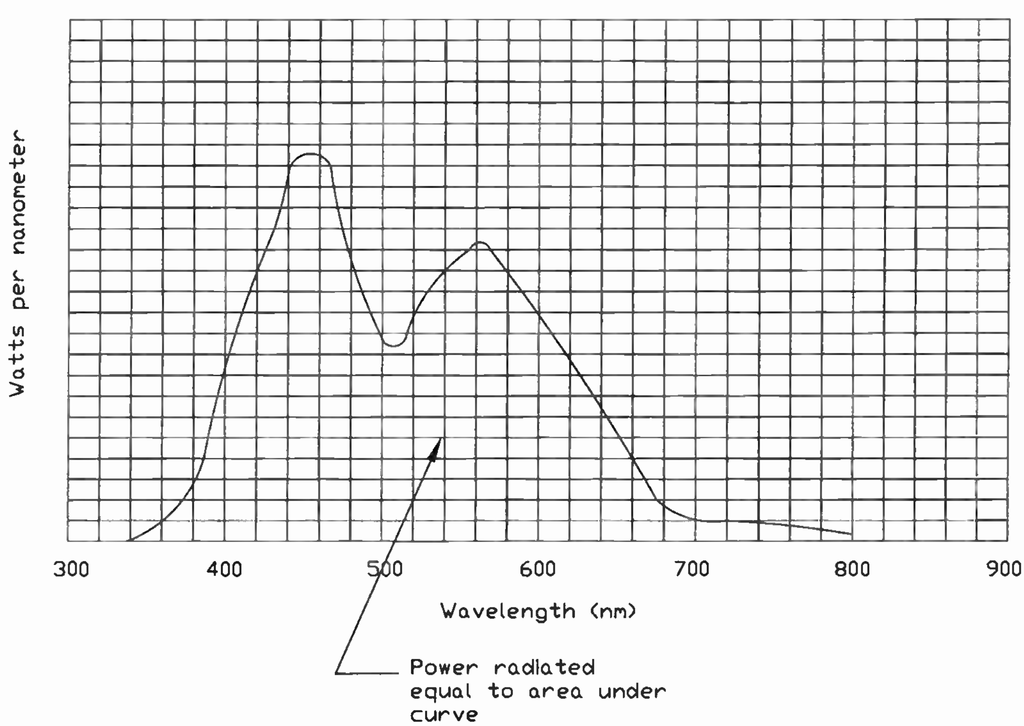


Figure 5.2-4. Power distribution of a monochrome video picture tube light source.[7]

MONOCHROME AND COLOR VISION

The color sensation associated with a light stimulus can be described in terms of three characteristics:

- Hue
- Saturation
- Brightness

The spectrum contains most of the principal hues: red, orange, yellow, green, blue and violet. Additional hues are obtained from mixtures of red and blue light—these constitute the purple colors. *Saturation* pertains to the strength of the hue. Spectrum colors are highly saturated. White and grays have no hue and, therefore, have zero saturation. Pastel colors are of low or intermediate saturation. *Brightness* pertains to the intensity of the stimulation. If a stimulus has high intensity, regardless of its hue, it is said to be bright.

The psychophysical analogs of hue, saturation, and brightness are:

- Dominant wavelength
- Excitation purity
- Luminance

This principle is illustrated in Table 5.2-1.

By means of definitions and standard response functions, which have received international acceptance through the International Commission on Illumination (CIE), the dominant wavelength, purity, and luminance of any stimulus of known spectral energy distribution may be determined by simple computations. Although roughly analogous to their psychophysical counterparts, the psychological attributes of hue, saturation, and brightness pertain to observer responses to light stimuli and are not subject to calculation. These sensation characteristics as applied to any given stimulus depend in part on other visual stimuli in the field of view and upon the immediately preceding stimulations.

Color sensations arise directly from the action of light on the eye. They are normally associated, however, with objects in the field of view from which the light comes. The objects themselves are therefore said to have color. *Object colors* may be described in terms of their hues and saturations, as is the case for light stimuli. The intensity aspect is usually referred to in terms of lightness, rather than brightness. The psychophysical analogs of lightness are *luminous reflectance* for reflecting objects and *luminous transmittance* for transmitting objects.

At low levels of illumination, objects may differ from one another in their lightness appearances but give rise to no sensation of hue or saturation. All objects appear to be of different shades of gray. Vision at low levels of illumination is called *scotopic vision,* as distinct from *photopic vision,* which takes place at higher levels of illumination. Table 5.2-2 compares the luminosity values for photopic and scotopic vision.

Only the rods of the retina are involved in scotopic vision; the cones play no part. As the fovea centralis is free of rods, scotopic vision takes place outside the fovea. Visual acuity of scotopic vision is low compared with photopic vision.

At high levels of illumination, where cone vision predominates, all vision is color vision. Reproducing systems such as black-and-white photography and monochrome video, however, cannot reproduce all three types of characteristics of colored objects. All images belong to the series of grays, differing only in their relative brightness.

The relative brightness of the reproduced image of any object depends primarily upon the luminance of the object as seen by the photographic or video camera. Depending upon the camera pickup element or the film, however, the dominant wavelength and purity of the light may also be of consequence. Most films and video pickup elements exhibit uniform sensitivity throughout the visible spectrum and, consequently,

Table 5.2-1
Psychophysical and psychological characteristics of color.

Psychophysical properties	Psychological properties
Dominant wavelength	Hue
Excitation purity	Saturation
Luminance	Brightness
Luminous transmittance	Lightness
Luminous reflectance	Lightness

Table 5.2-2
Relative luminosity values for photopic and scotopic vision.

Wavelength, nm	Photopic vision	Scotopic vision
390	0.00012	0.0022
400	0.0004	0.0093
410	0.0012	0.0348
420	0.0040	0.0966
430	0.0116	0.1998
440	0.023	0.3281
450	0.038	0.4550
460	0.060	0.5670
470	0.091	0.6760
480	0.139	0.7930
490	0.208	0.9040
500	0.323	0.9820
510	0.503	0.9970
520	0.710	0.9350
530	0.862	0.8110
540	0.954	0.6500
550	0.995	0.4810
560	0.995	0.3288
570	0.952	0.2076
580	0.870	0.1212
590	0.757	0.0655
600	0.631	0.0332
610	0.503	0.0159
620	0.381	0.0074
630	0.265	0.0033
640	0.175	0.0015
650	0.107	0.0007
660	0.061	0.0003
670	0.032	0.0001
680	0.017	0.0001
690	0.0082	
700	0.0041	
710	0.0021	
720	0.00105	
730	0.00052	
740	0.00025	
750	0.00012	
760	0.00006	

marked distortions in luminance as a function of dominant wavelength and purity are not encountered. Their spectral sensitivities seldom conform exactly to that of the human observer, however, so that some brightness distortions do exist.

Visual Requirements for Video

The objective in any type of visual reproduction system is to present to the viewer a combination of visual stimuli that can be readily interpreted as representing or having close association with a real viewing situation. It is by no means necessary that the light stimuli from the original scene be duplicated. There are certain characteristics in the reproduced image, however, that are necessary and others that are highly desirable. Only a general qualitative discussion of such characteristics will be given here.

In monochrome video, images of objects are distinguished from one another and from their backgrounds as a result of luminance differences. In order that detail in the picture be visible and that objects have clear, sharp edges, it is necessary that the video system be capable of rapid transitions from areas of one luminance level to another. This degree of resolution need not match that possible in the eye itself, but too low an effective resolution results in pictures with a fuzzy appearance and lacking fineness of detail.

Luminance range and the transfer characteristic associated with luminance reproduction are likewise of importance in monochrome video. Objects seen as white usually have minimum reflectances of approximately 80%. Black objects have reflectances of approximately 4%. This gives a luminance ratio of 20/1 in the range from white to black. To obtain the total luminance range in a scene, the reflectance range must be multiplied by the illumination range. In outdoor scenes, the illumination ratio between full sunlight and shadow may be as high as 100/1. The full luminance ranges involved with objects in such scenes cannot be reproduced in normal video reproduction equipment. Video systems must be capable of handling illumination ratios of at least 2, however, and ratios as high as 4 or 5 are desirable. This implies a luminance range on the output of the system of at least 40/1, with possible upper limits as high as 80/1 or 100/1.

Monochrome video transmits only luminance information, and the relative luminances of the images should correspond at least roughly to the relative luminances of the original objects. Red objects, for example, should not be reproduced markedly darker than objects of other hues but of the same luminance. Exact luminance reproduction, however, is by no means a necessity. Considerable distortion as a function of hue is acceptable in many applications. Luminance reproduction is probably of primary consequence only if detail in some hues becomes lost.

Images in monochrome video are transmitted one point, or small area, at a time. The complete picture image is repeatedly scanned at frequent intervals. If the frequency of scan is not sufficiently high, the picture appears to flicker. At frequencies above a *critical frequency* no flicker is apparent. The critical frequency changes as a function of luminance, being higher for higher luminance. The basic requirement for monochrome video is that the *field frequency* (the rate at which images are presented) be above the critical frequency for the highest image luminances.

Images of objects in color video are distinguished from one another by luminance differences or by differences in hue or saturation. Exact reproduction in the image of the original scene differences is not necessary or even attainable. Nevertheless, some reasonable correspondence must prevail because the luminance gradation requirements for color are essentially the same as those for monochrome video.

Luminous Considerations in Visual Response

Vision is considered in terms of physical, psychophysical and psychological quantities. The primary stimulus for vision is radiant energy. The study of this radiant energy in its various manifestations, including the effects on it of reflecting, refracting, and absorbing materials, is a study in physics. The response part of the visual process embodies the sensations and perceptions of seeing. Sensing and perceiving are mental operations and therefore belong to the field of psychology. Evaluation of radiant-energy stimuli in terms of the observer responses they evoke is within the realm of psychophysics. Because observer response sensations can be described only in terms of other sensations, psychophysical specifications of stimuli are made according to sensation equalities or differences.

Photometric Measurements

Evaluation of a radiant energy stimulus in terms of its brightness producing capacity is a photometric measurement. An instrument for making such measurements is called a *photometer.* In visual photometers, which must be used in obtaining basic photometric measurements, the two stimuli to be compared are normally directed into small adjacent parts of a viewing field. The stimulus to be evaluated is presented in the *test field;* the stimulus against which it is compared is presented in the *comparison field.* For most high-precision measurements the total size of the combined test and comparison fields is kept small, subtending about 2° at the eye. The area outside these fields is called the *surround.* Although the surround does not enter directly into the measurements, it has adaptation effects on the retina and thus affects the appearances of the test and comparison fields. It also influences the precision of measurement.

Luminosity Curve

A *luminosity curve* is a plot indicative of the relative brightnesses of spectrum colors of different wavelength or frequency. To a normal observer, the brightest part of a spectrum consisting of equal amounts of radiant flux per unit wavelength interval is at about 555 nm. Luminosity curves are therefore commonly normalized to have a value of *unity* at 555 nm. If, at some other wavelength, twice as much radiant flux as at 555 nm is required to obtain brightness equality with radiant flux at 555 nm, the luminosity at this wavelength is 0.5. The luminosity at any wavelength λ is, therefore, defined as the ratio P_{555}/P_{λ}, where P_{λ} denotes the amount of radiant flux at the wavelength λ, which is equal in brightness to a radiant flux of P_{555}.

The luminosity function that has been accepted as standard for photopic vision is given in Figure 5.2-5. Tabulated values at 10 nm intervals are given in Table 5.2-2. This function was agreed upon by the CIE in 1924. It is based upon considerable experimental work that was conducted over a number of years. Chief reliance in arriving at this function was based on the step-by-step equality-of-brightness method. Flicker photometry provided additional data.

In the scotopic range of intensities, the luminosity function is somewhat different from that of the photopic range. The two curves are compared in Figure 5.2-6. The two curves are similar in shape, but there is a shift for the scotopic curve of about 40 nm to the shorter wavelengths.

Measurements of luminosity in the scotopic range are usually made by the *threshold-of-vision* method. A single stimulus in a dark surround is used. The stimulus is presented to the observer at each of a number of different intensities, ranging from well below the threshold to intensities sufficiently high to be definitely visible. Determinations are made of the amount of energy, at each chosen wavelength, that is reported visible by the observer a certain percentage of the time, such as 50%. The reciprocal of this amount of energy determines the relative luminosity at the given wavelength. The wavelength plot is normalized to have a maximum value of 1.00 to give the scotopic luminosity function.

In the intensity region between scotopic and photopic vision, called the *Purkinje* or *mesopic region*, the measured luminosity function takes on sets of values intermediate between those obtained for scotopic and photopic vision. Relative luminosities of colors within the mesopic region will therefore vary, depending upon the particular intensity level at which the viewing takes place. Reds tend to become darker in approaching scotopic levels; greens and blues tend to become relatively lighter.

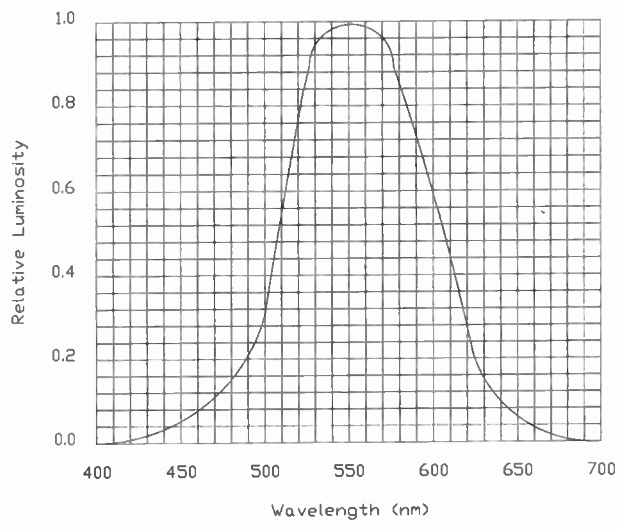


Figure 5.2-5. The photopic luminosity function.[7]

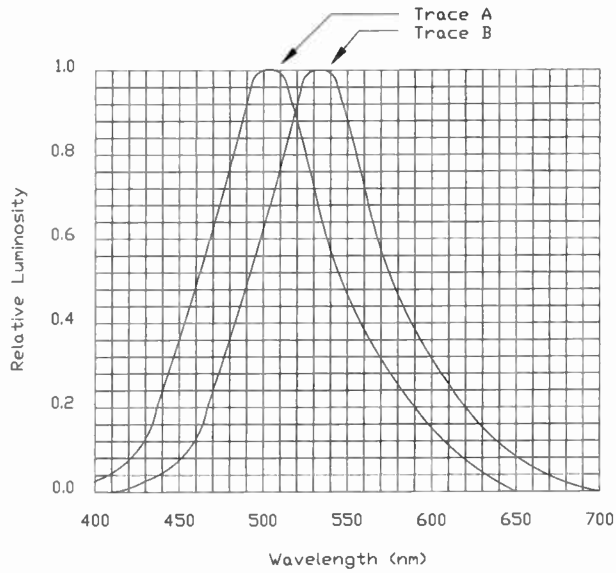


Figure 5.2-6. Scotopic luminosity function (trace *A*) as compared with photopic luminosity function (trace *B*).[7]

Luminance

Brightness is a term used to describe one of the characteristics of appearance of a source of radiant flux or of an object from which radiant flux is being reflected or transmitted. Brightness specifications of two or more sources of radiant flux should be indicative of their actual relative appearances. These appearances will depend in large part upon the viewing conditions, including the state of adaptation of the observer's eye.

Luminance, as indicated previously, is a psychophysical analog of brightness. It is subject to physical determination, independent of particular viewing and adaptation conditions. Because it is an analog of brightness, however, it is defined in such a way as to relate as closely as possible to brightness.

The luminosity function is the best established measure of the relative brightnesses of different spectral stimuli. In evaluating the luminance of a source of radiant flux consisting of many wavelengths of light, the amounts of radiant flux at the different wavelengths are weighted by the luminosity function, which converts radiant flux to luminous flux. As used in photometry, the term *luminance* is applied only to extended sources of light, not to point sources. For a given amount (and quality) of radiant flux reaching the eye, brightness will vary inversely with the effective area of the source.

Luminance is described in terms of luminous flux per unit projected area of the source. The greater the concentration of flux in the angle of view of a source, the brighter it appears. Luminance is therefore expressed in terms of amounts of flux per unit solid angle or *steradian*.

In considering the relative luminances of various objects of a scene to be captured and reproduced by a video system, it is convenient to normalize the luminance values so that the *white* in the region of principal illumination has a relative luminance value of 1.00. The relative luminance of any other object then becomes the ratio of its luminance to that of the white. This white is an object of highly diffusing surface with high and uniform reflectance throughout the visible spectrum. For purposes of computation, it may be idealized to have 100% reflectance and perfect diffusion.

Perception of Fine Detail

Detail is seen in an image because of brightness differences between small adjacent areas in a monochrome display or because of brightness, hue, or saturation differences in a color display. Visibility of detail in a picture is important because it determines the extent to which small or distant objects of a scene are visible, and because of its relationship to the *sharpness* appearance of the edges of objects.

Picture definition is probably the most acceptable term for describing the general characteristic of crispness, sharpness, or image-detail visibility in a picture. Picture definition depends upon characteristics of the eye, such as visual acuity, and upon a variety of characteristics of the picture-image medium, including its resolving power, luminance range, contrast, and image edge gradients.

Visual acuity may be measured in terms of the visual angle subtended by the smallest detail in an object that is visible. The *Landolt ring* is one type of test object frequently employed. The ring, which has a segment cut from it, is shown in any one of four orientations, with the opening at the top or bottom or on the right or left side. The observer identifies the location of this opening. The visual angle subtended by the opening that can be properly located 50% of the time is a measure of visual acuity.

Test object illuminance, contrast between the test object and its background, time of viewing, and other factors greatly affect visual acuity measurements. Up to a visual distance of about 20 ft (6 m) acuity is partially a function of distance, because of changes in shape of the eye lens in focusing. Beyond 20 ft it remains relatively constant. Visual acuity is highest for foveal vision, dropping off rapidly for retinal areas outside the fovea.

A black line on a light background is visible if it has a visual angle no greater than 0.5 s. This is not, however, a true measure of visual acuity. For visual acuity tests of the type described, normal vision, corresponding to a Snellen 20/20 rating, represents an angular discrimination of about 1 min. Separations between adjacent cones in the fovea and resolving power limitations of the eye lens give theoretical visual acuity values of about this same magnitude.

The extent to which a picture medium, such as a photographic or a video system, can reproduce fine detail is expressed in terms of *resolving power* or *resolution*. Resolution is a measure of the distance between two fine lines in the reproduced image that are visually distinct. The image is examined under the best possible conditions of viewing, including magnification.

Two types of test charts are commonly employed in determining resolving power, either a wedge of radial lines or groups of parallel lines at different pitches for each group. For either type of chart, the spaces between pairs of lines usually are made equal to the line widths. Figure 5.2-7 shows a test signal electronically generated by a video measuring test set.

Resolution in photography is usually expressed as the maximum number of lines (counting only the black ones or only the white ones) per millimeter that can be distinguished from one another. Measured values of resolving power depend upon a number of factors in addition to the photographic material itself. The most important of these typically are:

- Density differences between the black and the white lines of the test chart photographed
- Sharpness of focus of the test-chart image during exposure
- Contrast to which the photographic image is developed
- Composition of the developer

Sharpness of focus depends upon the general quality of the focusing lens, image and object distances from the lens, and the part of the projected field in which the image lies. In determining the resolving power of a photographic negative or positive material, a test chart is generally employed that has a high density difference, such as 3.0, between the black-and-white lines. A high quality lens is used, the projected field is limited and focusing is critically adjusted. Under these conditions, ordinary black-and-white photographic materials generally have resolving powers in the range of 30 to 200 line-pairs per mm. Special photographic materials are available with resolving powers greater than 1000 line-pairs per mm.

Resolution in a video system is expressed in terms of the maximum number of lines, counting both black and white, that are discernible in viewing a test chart. The value of horizontal (vertical lines) or vertical (horizontal lines) resolution is the number of lines equal to the vertical dimension of the raster. Vertical resolution in a well adjusted system equals the number of scanning lines, roughly 500 in conventional television. In normal broadcasting and reception practice, however, typical values of vertical resolution range from 350 to 400 lines. The theoretical limiting value for horizontal resolution (R_H) in a 525 line, 30 Hz frame rate system is given by:

$$R_H = \frac{2 \times 0.75 \times \Delta f}{30 \times 525}$$

$$R_H = 0.954 \times 10^{-4} \times \Delta f$$

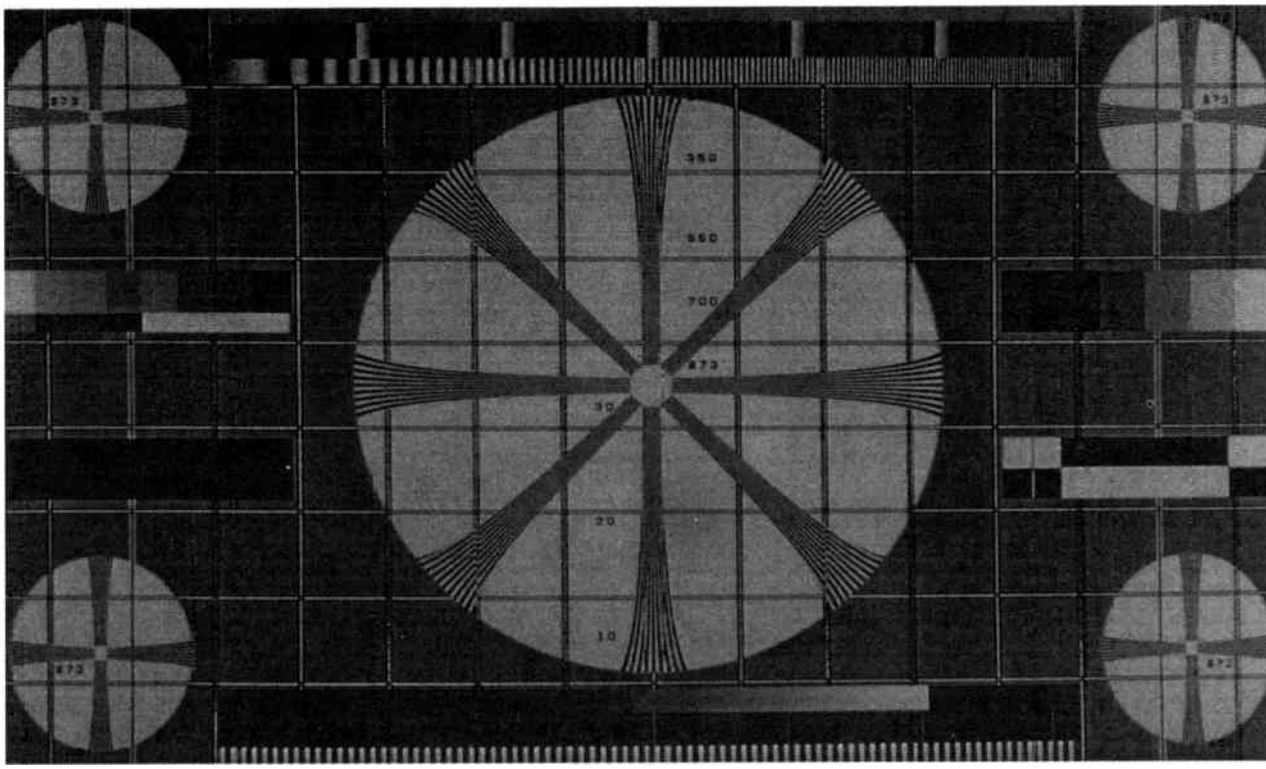


Figure 5.2-7. Test chart for high definition television applications produced by a signal waveform generator. The electronically produced pattern is used to check resolution, geometry, bandwidth, and color reproduction. (*Courtesy of Tektronix.*)

Where:

Δf = the available bandwidth frequency in Hz

The constants 30 and 525 represent the frame and line frequencies, respectively in the conventional United States television system. A factor of 2 is introduced because in one complete cycle both a black and a white line are obtainable. The factor 0.75 is necessary because of the receiver aspect ratio—the picture height is ¾ of the picture width. There is an additional reduction of about 15% (not included in the equation) in the theoretical value because of horizontal blanking time during which retrace takes place. A bandwidth of 4.25 MHz—typically that of the conventional terrestrial television system—thus makes possible a maximum resolution of about 345 lines.

Sharpness

The appearance evaluation of a picture image in terms of the edge characteristics of objects is called *sharpness.* The more clearly defined the line that separates dark areas from lighter ones, the greater the sharpness of the picture. Sharpness is, naturally, related to the transient curve in the image across an edge. The average gradient and the total density difference appear to be the most important characteristics. No physical measure has been devised, however, that in all cases will predict the sharpness (appearance) of an image.

Picture resolution and sharpness are to some extent interrelated but they are by no means perfectly correlated. Pictures ranked according to resolution measures may be rated somewhat differently on the basis of sharpness. Both resolution and sharpness are related to the more general characteristic of picture definition. For pictures in which, under particular viewing conditions, effective resolution is limited by the visual acuity of the eye rather than by picture resolution, sharpness is probably a good indication of picture definition. If visual acuity is not the limiting factor, however, picture definition depends to an appreciable extent on both resolution and sharpness.

HUMAN VISUAL SYSTEM

The human visual system (HVS) is powerful and exceeds the performance of artificial visual systems in almost all areas of comparison. Researchers have, therefore, studied the HVS extensively to ascertain the most efficient and effective methods of communicating information to the eye. An important component of this work has been the development of models of how humans see in an effort to improve image processing systems.

A Model for Image Quality

The classic approach to image quality assessment involves the presentation to a group of test subjects visual test material for evaluation and rating. The test material may include side-by-side display comparisons, or a variety of perception threshold presentations. One common visual comparison technique is called the *pair-comparison* method. A number of observers are asked to view a specified number of images at two or more distances. At each distance, the subjects are asked to rank the order of the images in terms of overall quality, clearness, and personal preference.

An image acquisition, storage, transmission and display system need not present more visual information to the viewer than the viewer can process. For this reason, image quality assessment is an important element in the development of any new video system. For example, in the development of a video compressing algorithm, the designer needs to know at what compression point impairments are visible to the average viewer.

Evaluation by human subjects, while an important part of this process, is also expensive and time consuming. Numerous efforts have been made to reduce the human visual system and its interaction with a display device to one or more mathematical models. (Some of this research is cited in [1–5].) The benefits of this research to design engineers is obvious: more timely feedback on new product offerings. In the development of advanced imaging systems, it is important to evaluate the visual performance of the system well in advance of the expensive fabrication process. Image capture and display design software, therefore, is valuable. Because image quality is the primary goal in many systems, tools that permit the engineer to analyze and emulate the captured and/or displayed image assist in the design process. Such tools provide early feedback on new display techniques and permit a wider range of prospective products to be evaluated.

After the system or algorithm has successfully passed the minimum criteria established by the model, it can be subjected to human evaluation. The model simulation requires the selection of many interrelated parameters. A series of experiments is typically conducted to improve the model to more closely approximate human visual perception.

HVS as it Affects Digital Video

A video compression system reduces the volume of data by exploiting spatial and temporal redundancies and by eliminating the data that cannot be displayed suitably by the associated display or imaging devices. [8] The main objective of compression is to retain as little data as possible to reproduce the original images without causing unacceptable distortion of the image. A compression system consists of the following components:

- *Digitization, sampling and segmentation.* Steps that convert analog signals on a specified grid of picture elements into digital representations, then divide the video input—first into frames, then into blocks.
- *Redundancy reduction.* The decorrelation of data into fewer useful data bits using certain invertible transformation techniques.
- *Entropy reduction.* The representation of digital data using fewer bits by dropping less significant information. This component causes distortion and it is the main contributor in *lossy* compression.
- *Entropy coding.* The assignment of code words (bit strings) of shorter length to more likely image symbols and code words of longer length to less likely symbols. This minimizes the average number of bits needed to code an image.

Key terms important to the understanding of this topic include the following:

- *Motion compensation:* The coding of video segments with consideration to their displacements in successive frames.
- *Spatial correlation:* The correlation of elements within a still image or a video frame for the purpose of bit-rate reduction.
- *Spectral correlation:* The correlation of different color components of image elements for the purpose of bit-rate reduction.
- *Temporal correlation:* The correlation between successive frames of a video file for the purpose of bit rate reduction.
- *Quantization compression:* The dropping of the less significant bits of image values to achieve higher compression.
- *Intraframe coding:* The encoding of a video frame by exploiting spatial redundancy within the frame.
- *Interframe coding:* The encoding of a frame by predicting its elements from elements of the previous frame.

The removal of spatial and temporal redundancies that exist in natural video imagery is essentially a lossless process. Given the correct techniques, an exact replica of the image can be reproduced at the viewing end of the system. Such lossless techniques are important for medical imaging applications and other demanding uses. These methods, however, usually realize only about a 2:1 compression efficiency. For video, a much higher compression ratio is required. Exploiting the inherent limitations of the human visual system can result in compression ratios of 50:1 or higher. These limitations include the following:

- Limited luminance response and very limited color response
- Reduced sensitivity to noise in high frequency components, such as at the edges of objects
- Reduced sensitivity to noise in brighter areas of the image

The goal of compression, then, is to discard all information in the image that is not absolutely necessary from the standpoint of what the HVS is capable of resolving. Such a system can be described as *psychovisually lossless.* Understanding the HVS, then, is fundamental to the development of advanced compression technologies that have made possible the ATSC digital television system.

REFERENCES

1. Grogan, Timothy A., "Image Evaluation with a Contour-Based Perceptual Model," *Human Vision, Visual Processing, and Digital Display III,* Bernice E. Rogowitz ed., Proc. SPIE 1666, pp. 188–197, 1992.
2. Barten, Peter G. J., "Physical Model for the Contrast Sensitivity of the Human Eye," *Human Vision, Visual Processing, and Digital Display III,* Bernice E. Rogowitz ed., Proc. SPIE 1666, pp. 57–72, 1992.
3. Daly, Scott, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity," *Human Vision, Visual Processing, and Digital Display III,* Bernice E. Rogowitz ed., Proc. SPIE 1666, pp. 2–15, 1992.
4. Reese, Greg, "Enhancing Images with Intensity-Dependent Spread Functions," *Human Vision, Visual Processing, and Digital Display III,* Bernice E. Rogowitz ed., Proc. SPIE 1666, pp. 253–261, 1992.
5. Martin, Russel A., Albert J. Ahumanda, Jr., and James O. Larimer, "Color Matrix Display Simulation Based Upon Luminance and Chromatic Contrast Sensitivity of Early Vision," *Human Vision, Visual Processing, and Digital Display III,* Bernice E. Rogowitz ed., Proc. SPIE 1666, pp. 336–342, 1992.
6. *IES Lighting Handbook,* Illuminating Engineering Society of North America, New York, 1981.
7. Fink, D. G., *Television Engineering,* 2nd ed., McGraw-Hill, New York, 1952.
8. Jerry C. Whitaker, "Audio and Video Compression, in *DTV: The Revolution in Electronic Imaging,* McGraw-Hill, New York, pp. 134–135, 1998.

BIBLIOGRAPHY

Benson, K. B., and J. C. Whitaker, *Television Engineering Handbook,* revised ed., McGraw-Hill, New York, 1991.

Boynton, R. M., *Human Color Vision,* Holt, New York, 1979.

Committee on Colorimetry, Optical Society of America, *The Science of Color,* New York, 1953.

Davson, H., *Physiology of the Eye,* 4th ed., Academic, New York, 1980.

Evans, R. M., W. T. Hanson, Jr., and W. L. Brewer, *Principles of Color Photography,* Wiley, New York, 1953.

Kingslake, R. (ed.), *Applied Optics and Optical Engineering,* Vol. 1, Academic, New York, 1965.

Polysak, S. L., *The Retina,* University of Chicago Press, 1941.

Richards, C. J., *Electronic Displays and Data Systems: Constructional Practice,* McGraw-Hill, New York, 1973.

Schade, O. H., "Electro-optical Characteristics of Television Systems," *RCA Review,* Vol. 9, pp. 5–37, 245–286, 490–530, 653–686, 1948.

Whitaker, J. C., *Electronic Displays: Technology, Design and Applications,* McGraw-Hill, New York, 1993.

Wright, W. D., *Researches on Normal and Defective Colour Vision,* Mosby, St. Louis, 1947.

Wright, W. D., *The Measurement of Colour,* 4th ed., Adam Hilger, London, 1969.

DEFINING TERMS

Brightness. A term used to describe one of the characteristics of appearance of a source of radiant flux or of an object from which radiant flux is being reflected or transmitted.

Critical frequency. The rate of picture presentation, as in a video system or motion picture display, above which the presented image ceases to give the appearance of flickering. The critical frequency changes as a function of luminance, being higher for higher luminance.

Dispersion. The variation of refraction as a function of wavelength.

Energy distribution curve. A plot of the power distribution of a source of light giving the energy radiated at each wavelength per nanometer of wavelength.

Field frequency. The rate at which images in an electronic imaging system are presented. In conventional terrestrial television (NTSC), two fields are presented each second to make up one *frame* (a complete picture).

Refraction. The bending of light as it passes from one medium to another, such as from air to glass, at an angle.

Resolution. The extent to which an imaging system, such as a photographic or a video system, can reproduce fine detail.

Sharpness. The appearance evaluation of a picture image in terms of the edge characteristics of objects contained therein. The more clearly defined the lines separating dark areas from light ones, the greater the sharpness of the picture.

FOR FURTHER INFORMATION

The International Society for Optical Engineering (SPIE) (Bellingham, Washington) offers a number of publications examining the characteristics of the human visual system. The organization also conducts technical seminars on various topics relating to optics and the application of optical technologies.

The following books are recommended for further reading:

Jerry Whitaker, *Electronic Displays: Technology, Design and Application,* McGraw-Hill, New York, 1993.

5.3
VIDEO PICKUP DEVICES AND SYSTEMS

PETER GLOEGGLER
SONY CORPORATION, PARK RIDGE, NJ

INTRODUCTION

Color television, color printing and color photography all share the *tristimulus* concept of reproducing a colored image. In each of these systems, a combination of basic colors is used to create the wide spectrum of colors each of these mediums is capable of producing. Figure 5.3-1 shows a plot of the three NTSC color primaries on the CIE chromaticity diagram. The full range of colors contained within the red, green and blue (RGB) triangle on this diagram can be reproduced by the NTSC color television system. In its most basic form, all of the color television systems currently in use separate the incoming light into three separate images, corresponding to the red, green and blue spectra of the original scene.

Solid-state imagers are most commonly used to convert the three optical images into three electrical video signals that correspond to these images. These signals are then transmitted to another location or to the home for display. The signals are frequently combined into a single analog composite signal, using NTSC, *Phase Alternation Line* (PAL) or *Sequential Couleur Avec Memoire* (SECAM) encoding before transmission. At the receiver (or monitor), these composite signals are separated again into the red, green, and blue components and applied to a color display device, most commonly a *cathode ray tube* (CRT).

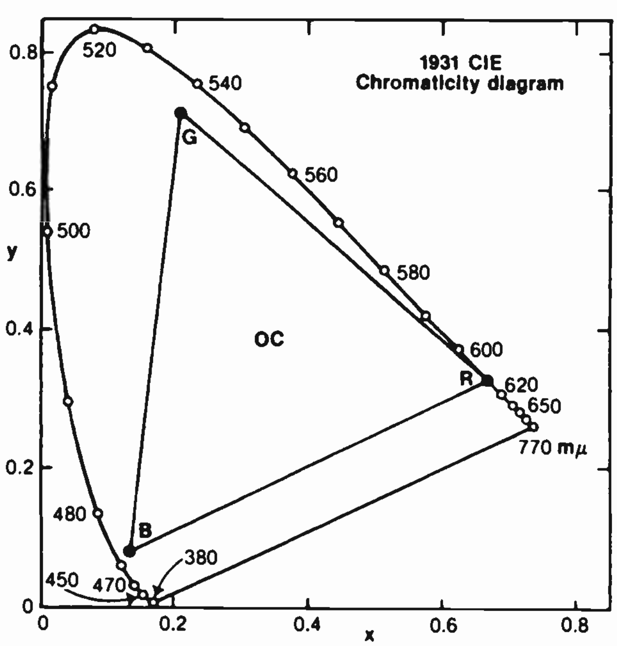


Figure 5.3-1. CIE chromaticity diagram showing the NTSC color primaries. *(Courtesy of Wiley Publishing.)*

In the United States, the existing composite analog transmission to the home using the NTSC encoding format will be supplemented and eventually replaced by the new component digital transmission system (the three color components are transmitted separately) as recommended by the Advanced Television Standards Committee (ATSC). With the FCC's acceptance of this recommendation, the migration to digital television is underway in the United States. Unlike NTSC, where a single picture format is mandated, the new standard for digital television describes a family of acceptable picture formats. (See Table 3 of the ATSC Document A/53 for a listing of the permissible transmission formats.)

In the case of the CRT, an array of red, green and blue phosphors on the screen are excited by electron beams corresponding to the red, green and blue signals to create the final color image.

THE CAMERA SYSTEM

Television became a practical reality with the invention of two electronic devices that use an electron beam in a vacuum envelope to overcome the limitations of the electromechanical devices used in earlier experiments with television. Both devices use an electron beam, albeit to opposite purposes. The first device, the CRT, uses an electron beam to scan a phosphor layer that emits light in direct proportion to the intensity of the electron beam striking it. Because the intensity of the electron beam can—in turn—be controlled by an electrical signal applied to the CRT, this electrical signal can be used to create an optical image.

The second device, the *image pickup tube,* also uses an electron beam. However, in this case, the electron beam scans an image focused on the photosensitive layer of the tube and produces an electrical signal that is proportional to the optical image. The image pickup tube and its solid-state successor, the *charge coupled device* (CCD) imager, constitute the heart of the television camera.

The video camera is only the starting point in bringing a television image to the viewer, but it is one of the most exciting and challenging parts of the system. Creating a modern video camera requires the latest

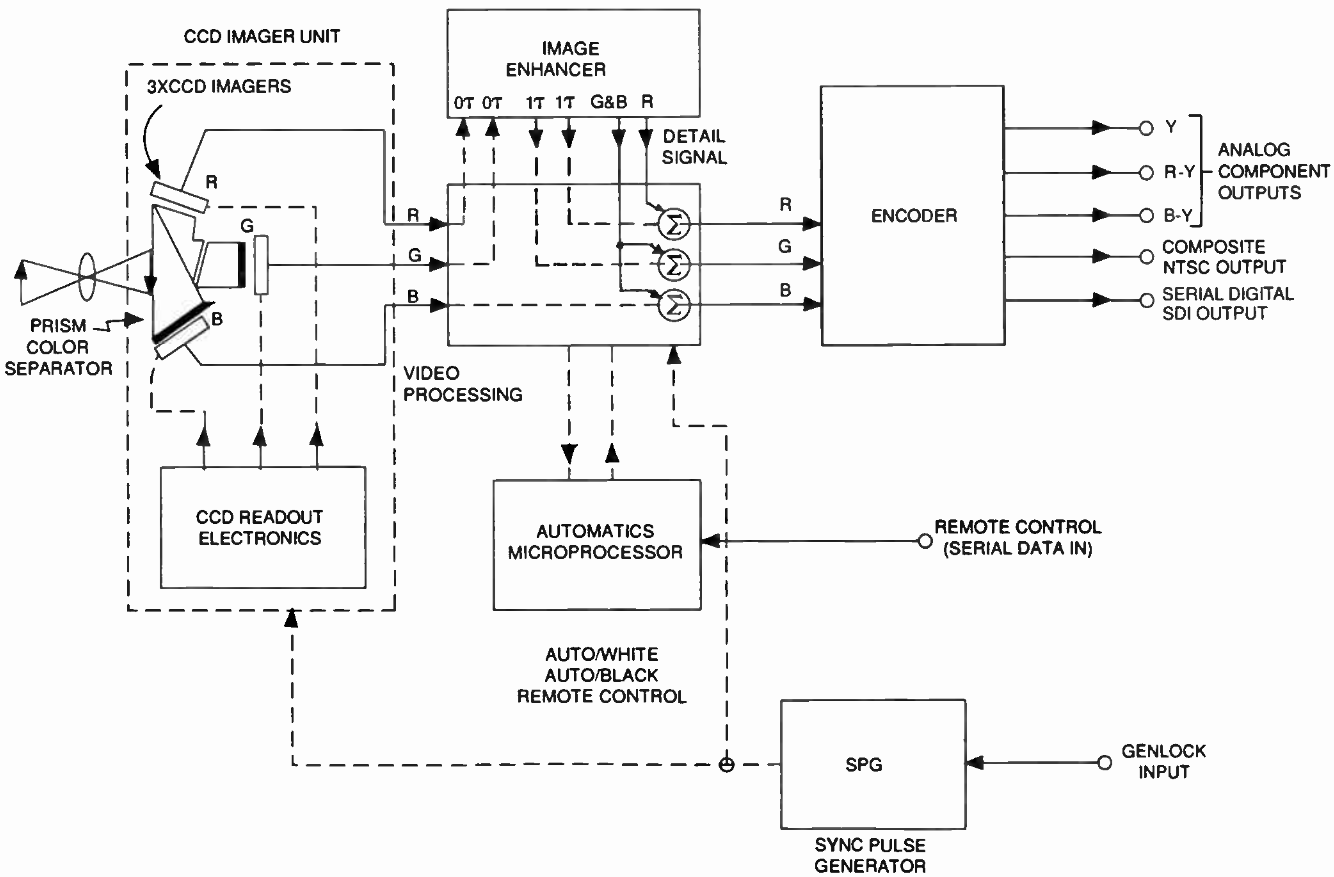


Figure 5.3-2. Basic block diagram of a video camera.

technology from a wide variety of disciplines: optics, integrated circuit technology, microprocessor control and manufacturing technology.

Video cameras, from the largest units used in the studio, to small cameras used for *electronic news gathering* (ENG), share many common elements (see Figure 5.3-2). Cameras, typically contain these major parts:

- A lens to capture the scene
- Color separation optics to separate the incoming light into three primary color images: a red, green and blue image
- Three pickup devices that convert the red, green and blue images into three electrical signals, the RGB video signals
- Video processing circuitry that performs a variety of processing steps to:
  correct for errors in the lens and optical system, correct for the color temperature of the scene illumination, compress the dynamic range, increase the apparent sharpness of the image through the use of image enhancement techniques, and introduce gamma pre-correction
- Encoding system—circuitry that encodes the basic RGB signals created by the camera into analog component, analog composite or digital signals appropriate for recording or distribution
- Timing signal generator—circuitry that produces the waveforms necessary for the internal operation of the camera
- Power supply—circuitry that generates the voltages required to operate the various parts of the camera.

Camera System Details

To illustrate the complex processes that occur in a video camera, we will follow the rays of light entering the lens, conversion of the light to an electrical signal and the processing of the signal to the final output signal delivered by the camera.

The Optical System

The optical system of the camera is used to form a precise image of the original scene on the surface of the imaging devices. The optical system consists of:

- A lens to capture the image
- Optical filters to condition the image
- The color separation system to split the incoming light into the three primary color components.

Clearly, the quality of the optical system is critical to the performance of the overall camera. In the world of optics, there is usually a close correlation between cost and performance. It is not surprising, therefore, to find a somewhat lower level of performance in the optical system of lower cost cameras and lenses.

Brief descriptions of the optical components used in a camera are given here. For further information, see references 1 and 3.

The Lens

With the exception of the highest levels of program production, where fixed focal length (*prime lenses*) are sometimes used, the *zoom lens* is the universal lens used with virtually all television cameras. Zoom lenses are available at a wide range of prices and performance levels, up to and including performance levels required for use with high-definition television (HDTV) systems.

At first look, the requirements for a lens intended for use with a TV camera appear to be quite similar to lenses intended for use with a film camera, and indeed, film lenses are routinely used with black and white and single imager color video cameras. Unfortunately, lenses appropriate for a high quality video work differ in one critical parameter from lenses designed for film cameras. The *back focal distance* (the distance from the end of the lens to the plane where the image is formed) is increased significantly compared to film lenses to allow the insertion of the prism type color separation system between the lens and the imagers. For this reason, it is not possible to use the immense installed base of high quality film lenses directly with a professional (prism color separation) color video camera. A sophisticated optical adapter is necessary to use film lenses directly. Such *relay optic* adapters are available and are typically used with extreme wide angle and extreme telephoto film lenses for shooting requirements that cannot be achieved with available video zoom lenses.

With the FCC's acceptance of a new digital broadcast system in 1997, a new factor has been added to the television picture. Broadcasting in a widescreen 16:9 aspect ratio as compared to the current 4:3 ratio of NTSC transmission will be permitted on this digital channel. Video cameras are now being made available with imagers capable of producing a wide aspect ratio image. The decision has been made that the image diagonal dimension for cameras that produce 16:9 aspect ratio pictures and cameras that produce the current 4:3 pictures shall be the same. This means that a 16:9 camera can use the same lens that is used with a 4:3 aspect ratio camera. On the other hand the 16:9 aspect ratio and the higher picture quality that can be delivered to the home by the new new digital broadcast system, mandates the use of the higher quality offerings of lens suppliers to fully support these new picture quality levels (for more details on lens parameters see Appendix B).

Optical Filters

Several optical filters are used to achieve the high level of performance found in a modern camera:

- Color correction and neutral density filters
- Infrared filter
- Quarter wavelength filter
- Anti-aliasing filter.

Color Correction and Neutral Density Filters

The human eye adapts to changes in the characteristics of the surrounding illumination. Thus, after a short period of adaptation, the color of an object will be perceived by the eye as the same in daylight or in incandescent illumination. Cameras, whether they are film or electronic, do not adapt to the illumination, and so the color of an object will differ with the color temperature of the illuminant if appropriate adjustments are not made in the camera. Color cameras are designed to operate with lighting of a nominal color temperature, most commonly 3,200° K—the lighting provided by standard halogen studio lights. (See Appendix C for a more detailed discussion of the concept of color temperature.)

However, the camera must have the ability to operate properly with various other illumination sources. A combination of optical color correction filters and electrical adjustments of the relative gain of the camera's red, green and blue processing channels are used to reproduce the object correctly under different illumination. Outdoor scenes challenge the camera with a wide range of illumination, rich in reds early in the day and late in the afternoon, and dominated by blue at midday. In terms of color temperature, the video camera must be able to achieve white balance for illumination with a color temperature of less than 3,000° K for warm incandescent lighting, to color temperatures as high as 10,000° K on a bright, wintry day outdoors. Outdoor illumination is frequently much brighter than indoor lighting, and for this reason a neutral density filter is commonly used in conjunction with a daylight correction filter to keep the lens iris at the most favorable portion of its operating range.

Color cameras commonly include a four position filter wheel that allows convenient selection of the most appropriate combination of neutral density and color correction filters for a given scene illumination. Although many cameras have the ability to electronically achieve white balance even when an incorrect filter is used, the signal-to-noise (S/N) and the ability to handle pictures with very high contrast will be impaired.

Infrared Filter

The response of a CCD imager extends beyond the visible range to include the near infrared region. To avoid incorrect color reproduction from the unwanted response of the CCD or pickup tube imager to infrared light (wavelengths below the visible spectrum), video cameras normally include an infrared filter in the optical path to limit the response of the camera to the visible spectrum.

Quarter Wavelength Filter

The dichroic coatings used in the prism type color separation system respond differently to light of different polarization. To avoid a change in colorimetry with a change in polarization in the scene, a quarter wavelength filter is introduced before the prism in high performance cameras. The quarter wavelength filter alters the plane polarization of the incoming light to circular polarization. Lower cost cameras may not include this filter due to its high cost.

Anti-Aliasing Filter

A filter relatively new to the design of video cameras, the anti-aliasing filter has the difficult task of removing the extremely fine picture detail that can produce objectionable aliasing in a solid-state imager but have relatively little affect on picture detail within the resolving power of the imager. The quality of this filter is the key element that controls the compromise between useful resolving power and curtailment of the aliasing phenomenon in the imager.

Different from the smooth homogeneous imaging layer used in the pickup tube, the CCD solid-state imager uses an orthogonal, two-dimensional array of separate discrete sensors. Each sensor element develops a charge that is directly proportional to the light level impinging on this specific sensor element. The CCD array is, therefore, a collection of discrete electronic samples of the original optical image. Recognizing the CCD imager as a sampled data system, it becomes clear that the Nyquist sampling theorem applies. That is, errors will result when an input is sampled at a rate that is less than twice the highest frequency in the input. Here, the frequencies are spatial rather than termporal, but the results are the spatial equivalent.

To avoid the aliasing phenomenon in the electrical output of the imager, an optical low-pass filter must be added to the camera to remove fine image detail above the spatial Nyquist limit. The optical low-pass filter is similar in design to a multielement *star filter*, and one of the signatures of this optical filter—found in almost all CCD cameras—is the star filter-like rays emanating from a strong highlight. Two-axis optical low-pass filters are most frequently used in industrial grade CCD cameras, while three-axis filters are normally provided in broadcast grade cameras, because aliasing tends to be much lower with the more complex three-axis filter.

The Color Separation System

A color camera requires an optical system that separates the incoming image into its three component colors. The type and quality of the color separation system used determines not only the faithfulness with which the camera reproduces colors, but also determines the physical arrangement of the imagers and thereby the mechanical design of the camera.

Current color cameras use one of the following designs:

- A single imager camera with color separation achieved by color filters positioned over the individual pixel
- Two imager camera, with one imager for luminance and the other for color reproduction
- Three imager camera, with prism type optics for color separation.

Sensitivity and the ability to accurately reproduce colors, particularly when luminance values are close to black or peak white, are compromised in the two imager cameras. Resolution, sensitivity and color reproduction near black and peak white luminance values are compromised in the one imager camera. For this reason, applications of one and two imager cameras tend to be limited to consumer and industrial applications, while prism type cameras are usually the choice for broadcast and program production.

For single and two imager type cameras, two different methods of color separation are available: *the RGB Primary Color Separation System* and complementary type color separation system. The RGB primary color system uses red, green and blue optical filters in front of the CCD devices. The *complementary type* color separation system uses yellow, cyan and magenta optical filters in front of the CCDs.

The *prism type color separation system* has become the dominant technique for high-end cameras by virtue of its compactness and overall performance (see Table 5.3-1). (An illustration of the three types of color separation systems and a more detailed description of the prism type system is provided in Appendix 1, *Color Separation Systems.*)

Imaging Devices

The imaging device is the transducer that converts the optical image into an electrical analog of the image. Cameras used in television broadcasting are almost entirely color cameras and use one, two or three imaging devices.

Image Diagonal

Image size is one of the more important factors that determine the overall performance of a camera. High performance cameras suitable for broadcast and production commonly use 2/3 in. or 1/2 in. format CCD imagers. The 2/3 in. image format has established itself as the optimum compromise between camera size, weight and performance. Limitations on minimum line width and other limitations of integrated circuit technology make it difficult to shrink the size of the imager and maintain the same level of performance. For a given type of CCD imager, 2/3 in. format cameras, therefore, tend to offer a higher level of performance than cameras using 1/2 in. imagers. In addition, a theoretical analysis indicates that the design requirements for the lens and the optical system become more severe as the image diagonal is reduced.

Sensitivity is one more specification of the camera that is strongly related to image diagonal. The prism color separation system has a basic limitation on the rate of convergence of the light rays that pass through it (the rate of convergence is limited to $f/1.4$ in the commonly used *three-piece* prism system) and as a result, it is possible to use a lens with almost twice the light gathering power (larger objective diameter)

Table 5.3-1
Relative performance of color separation systems
(Performance in a camera using 2/3 in., f/5.6, 768 pixel CCDs).

Color Separation System	Number of Imagers Required	Resolution	Sensitivity at 2000 Lux (Full Output)	Color Fidelity	Cost
Three Imager Cameras: Prism Type	3	Good, −700 TVL	Good, −f/5.6	Good	High
Two Imager Cameras: Luma/Chroma Type	2	Good, −560 TVL	Fair, −f/4.0	Fair*	Medium
Single Imager Cameras: Color Stripe: RGB Primary	1	Poor, −330 TVL	Poor, −f/2.8	Fair*	Low
Color Stripe: Complementary	1	Fair, −330 TVL	Fair, −f/4.0	Poor*	Low
Mosaic Type: Primary Color	1	Fair, −440 TVL	Poor, −f/2.8	Fair*	Low
Mosaic Type: Complementary	1	Fair, −440 TVL	Fair, −f/4.0?	Poor*	Low

* The color stripe/mosaic method of color separation tends to be ineffective for very dark colors (luminance <5 IRE) and very bright colors (luminance >95 IRE).

with a 2/3 in. format camera as compared to a 1/2 in. format camera. As a negative, the cost of the imagers and the overall size of the camera tends to increase when CCDs with larger image diagonal are used. It is not surprising to find that 2/3 in. imagers are the choice for the highest cost cameras, while 1/2 in. imagers predominate among the lower cost cameras offered by manufacturers.

The Camera Pickup Tube

In 1970, Willard Boll and George Smith of the Bell Labs announced the discovery of the CCD and with this development, the dream of a solid-state imager for use in a television camera took a big step toward reality. After almost 30 years of development, the CCD imager is displacing the pickup tube in virtually all cameras. This chapter, therefore, concentrates on the details of the new solid-state imagers and limits the discussion of cameras using pickup tubes.

The pickup tube was one of the key inventions that made the television system possible. In the pickup tube, the optical image focused on the photosensitive layer is scanned by an electron beam and an electrical signal directly proportional to the optical image is provided at the output of the device.

In a series of steps, the early pickup tubes evolved from the *Image Iconoscope*, demonstrated by Zworykin in 1933 and frequently described as the most important invention in the development of television cameras, to the *Image Orthicon, Vidicon*, and finally to the *Plumbicon*® and *Saticon*® pickup tubes. Refined over a period of many years, the Plumbicon and Saticon achieved a level of performance that allowed these two devices to reign as the unchallenged imagers for high performance cameras for a period of approximately 20 years. Of the two devices, the Plumbicon is considered to have somewhat more desirable characteristics overall. Short- and long-term image retention, in particular, tend to be better and, for this reason, the Plumbicon became the most commonly used imager in broadcast cameras. Saticons were easier to manufacture than Plumbicons and as a result were more commonly found in lower cost cameras.

The CCD Solid-State Imager

Ever since the introduction of the first professional CCD camera by RCA in 1983, the imagination of the broadcaster has been sparked by the concept of a new kind of camera, one that would eliminate the intensive care and maintenance required to sustain the performance of a pickup tube camera.

Another three years would pass before the dream would become at least a partial reality. In 1986, a new solid-state camera, the Sony BVP-5 camera was introduced with a docking videotape recorder (VTR) for ENG use. The BVP-5 went on to become the first CCD camera with widespread use in a broadcast environment.

Gone was the notion of the perfect camera; the considerable benefits of the solid-state imager were tempered by a whole new array of terms to describe the imperfections and foibles of these new devices. The terms frame transfer, interline transfer, frame interline transfer, pixels, aliasing, vertical smear, fixed pattern noise, spatial offset and many others entered the technical vocabulary and needed to be understood to decipher the competing claims of this new technology.

It is useful to compare the capability of these devices against the well known characteristics of the Plumbicon and Saticon pickup tubes, as a frame of reference. The advantages of CCD imagers compared to pickup tubes include the following:

- Cameras using solid-state imagers are smaller, lighter and consume less power than cameras of comparable performance using pick up tubes
- CCD imagers do not change characteristics or wear out with use

- CCD cameras are not damaged by strong highlights or high ambient temperatures
- Short- and long-term image burn-in and *comet tailing,* objectionable in pickup tubes, are more easily avoided
- CCD camera resolution and registration accuracy are uniform over the whole screen, compared to significant fall-off in certain areas with pickup tubes
- No centering or registration adjustments are required with CCD cameras
- A whole new concept, the *electronic shutter* to reduce blur in fast motion is possible with the CCD imager
- *Dynamic resolution,* (resolution with the subject or the camera in motion) is significantly better with the CCD imager than with the pickup tube.

The disadvantages of CCD imagers compared to pickup tubes include the following:

- *Aliasing,* caused by the limited number of discrete picture elements in the CCD, can cause objectionable artifacts with the presence of very fine details in the picture
- Diagonal lines have a stepping effect because of the limited number of picture elements in the CCD
- The dark noise of one picture element may differ slightly from the noise level of the next pixel, resulting in an objectionable phenomenon called *fixed pattern noise* when gain boost is used
- Registration of the imagers is fixed in a CCD camera. Longitudinal and lateral chromatic aberrations in the lens optical system are much more difficult to correct in CCD cameras
- The *vertical smear* phenomenon, a line from top to bottom when an excessive highlight is encountered, is only totally eliminated in the frame transfer type CCD imager
- *Dynamic picture distortion* is an artifact that introduces a bowing effect in a vertical structure as the camera is panned quickly.

The Basics of the CCD Imager

For all the wonders promised by this new solid-state era, it must be remembered that the CCD is, first and foremost, an analog device, just like the pickup tube. It, thus, possesses a wide range of analog characteristics, some labeled with the same familiar terms used to describe the characteristics of pickup tubes, others labeled with terminology specific to this new technology.

The CCD is radically different in concept from its photoconductive counterparts. In place of the amorphous layer in the pickup tube that is scanned by an electron beam, the CCD consists of a large number of discrete picture elements (pixels) arranged in an orthogonal, two-dimensional array. Each pixel accumulates an electronic charge directly proportional to the unique light level stimulating it. The CCD array, then, is a large collection of discrete electronic samples of the optical image.

Unlike the smooth, continuous output of the pickup tube as it is scanned by the electron beam, the output signal of the CCD imager is a series of discrete samples, generated as the information contained in the array is read, pixel by pixel. The resulting output signal of the CCD imager is very much like the analog output signal of a pickup tube if it were sampled at a corresponding rate.

Similarly, the CCD behaves as other sampled data systems. For example, the Nyquist sampling theorem mandates that the input to a sampled data system must be bandwidth limited at one half the sampling frequency to avoid aliasing errors in the sampled data. The nature of the aliasing experienced with a solid state imager and the need for an anti-aliasing filter may be analyzed on this basis.

In the solid-state imager, the sampling takes place in the optical domain, and for that reason, the bandwidth limitation prescribed by the Nyquist theorem must also take place in the optical domain. This requires use of an optical filter that will attenuate the high frequency energy impinging on the imager that can cause aliasing, without attenuating the useful in-band frequencies. Difficult trade-offs must be considered in the design of this filter. A more effective optical filter reduces aliasing, but also reduces the ability to reproduce fine detail in the picture.

Figure 5.3-3 shows the theoretical output of a CCD imager that has the form of Sin X/X, the response of the optical prefilter to reduce aliasing and the typical response of a Plumbicon pickup tube for comparison.

The Nyquist theorem now also makes it clear that an increase in the number of pixels in the imager is equivalent to an increase in the sampling frequency in a sampled data system. An increase in the resolution of the overall system is thereby possible.

The concept of sampling also explains the number of horizontal elements in current CCDs, which feature 768 horizontal pixels. A quick calculation shows that this number of 768 samples per active horizontal line corresponds to four samples per color subcarrier cycle, better known as $4f_{sc}$ sampling:

$$4 \times f_{sc} = 4 \times 3.579545 \text{ MHz} = 14.318 \text{ MHz}$$

and

$$14.318 \times 10^{6} \frac{\text{samples}}{\text{sec}} \times 53.6 \times 10^{-6} \frac{\text{sec}}{\text{activeline}} = 768 \frac{\text{samples}}{\text{line}}$$

Increasing the number of pixels in the CCD imager has the potential to increase the resolution and to lower the aliasing phenomenon. However, technical difficulties arise when the sampling frequency is not a multiple of the subcarrier frequency. For this reason, the commonly used 420,000 pixel CCD with 768 horizontal pixels is now being replaced with a higher resolution 520,000 pixel CCD with 960 horizontal pixel and an 18 MHz clock, effectively providing 5 samples per subcarrier cycle.

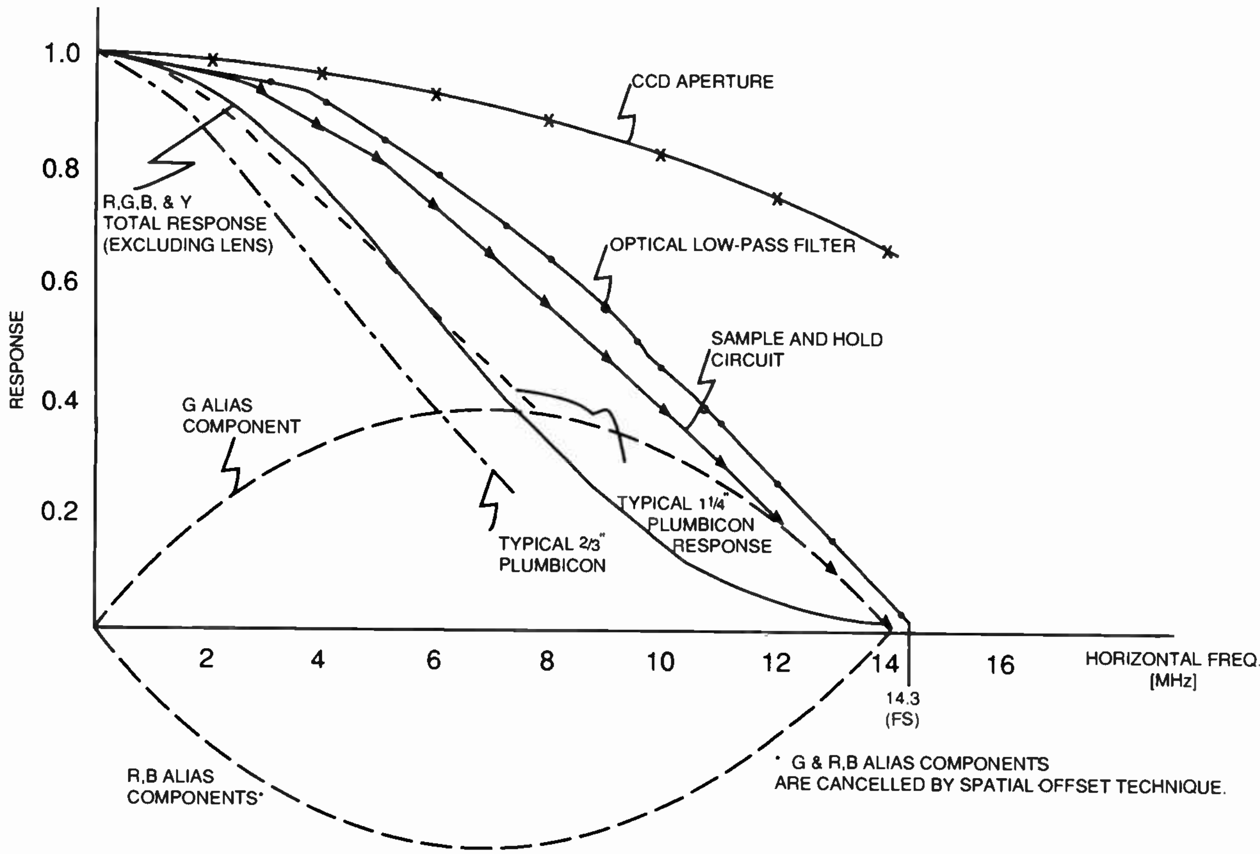


Figure 5.3-3. Typical MTF of a 3-chip CCD camera compared with the typical responses of Plumbicon tubes. Alias components in the G and the R and B signals cancel each other as a result of spatial offset.

The method for transforming the samples from the CCD also affect performance.

Frame Transfer CCD

The *frame transfer CCD* array has two separate sections: an array of imaging cells to capture the image, and an equally large array of cells carefully protected from light to store the image. A means of transferring the charges from the imaging section to the storage sections is also needed.

The basic diagram of such a device is shown in Figure 5.3-4. During the active field period, the cells of the imaging sensor array accumulate charges representative of the optical image focused on it. The information in the sensor array is shifted to the lower storage array at high speed during the next vertical interval. The now empty imager array will be charged with the succeeding image during the next active field period while, at the same time, picture information in the storage section is read to create the camera output signal.

An important characteristic of the *frame transfer* CCD (FT CCD) is its optical efficiency. Each picture element is large and virtually contiguous with the next element to capture virtually all of the incoming light.

The FT CCD also offers an efficient mechanism to transfer the charges from the imaging section to the storage section. The imaging elements function both as imagers and as effective shift registers to transfer the contents of the imager section into the storage section.

The weakness in the basic FT imager is the contamination that can result as the charges are shifted from the imaging array to the storage array. The sensors are

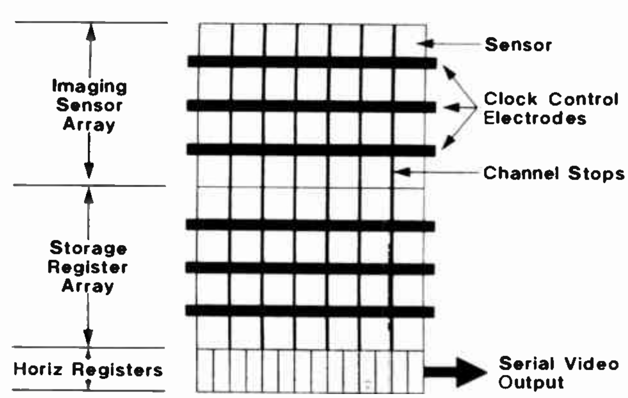


Figure 5.3-4. Frame transfer CCD structure.

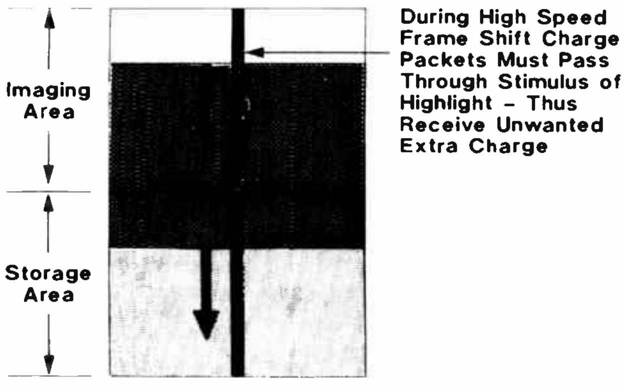


Figure 5.3-5. Mechanism of vertical smear in the frame transfer CCD.

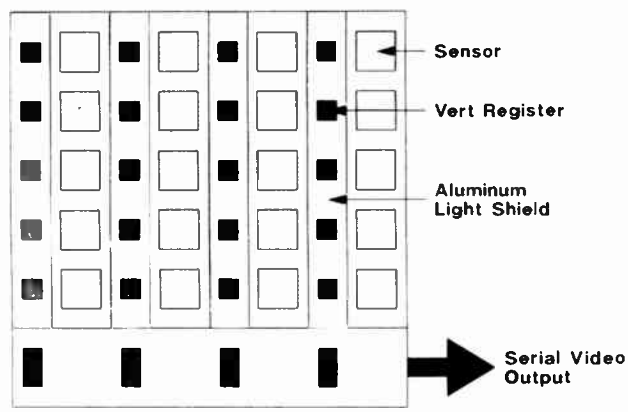


Figure 5.3-7. Structure of the interline CCD.

still being stimulated by the continuously impinging optical image during the short time the content of the imager section is transferred to the storage section. Erroneous signals at levels approximately 50 dB below normal signal level may result, levels clearly not acceptable in a professional camera. The name vertical smear has been applied to describe this phenomenon (see Figure 5.3-5).

A mechanical shutter interposed between the lens and the prism provides a solution to this problem. The shutter is synchronized with the vertical blanking interval to totally block the incoming light during the transfer process and thereby avoid the contamination that can otherwise occur (see Figure 5.3-6). The use of a FT CCD imager with a mechanical shutter form a viable imaging system for a broadcast camera. The FT CCD was introduced to the broadcast industry by RCA in 1983 and continues to be used in current broadcast cameras.

Interline-Transfer (IT) CCD

A completely different approach to the problem of moving the charge packets from the imaging array to the storage array is taken in the *interline transfer* (IT CCD). Storage elements protected from light are located next to each of the imaging elements. A simplified structure of the IT CCD is shown in Figure 5.3-7.

During the active field period, the sensor array again accumulates charges proportional to the image focused on it. However, during the next vertical interval, these charges are quickly transferred to the adjacent light protected storage elements, called the *interline storage register*. During the succeeding active field, the now empty sensor array charges with the next image, while the previous image is read from the interline storage register to become the video output of the camera. Comparing the IT CCD to the FT CCD imager, note that a part of optical sensing area had to be sacrificed in the IT CCD to make room for the interline storage

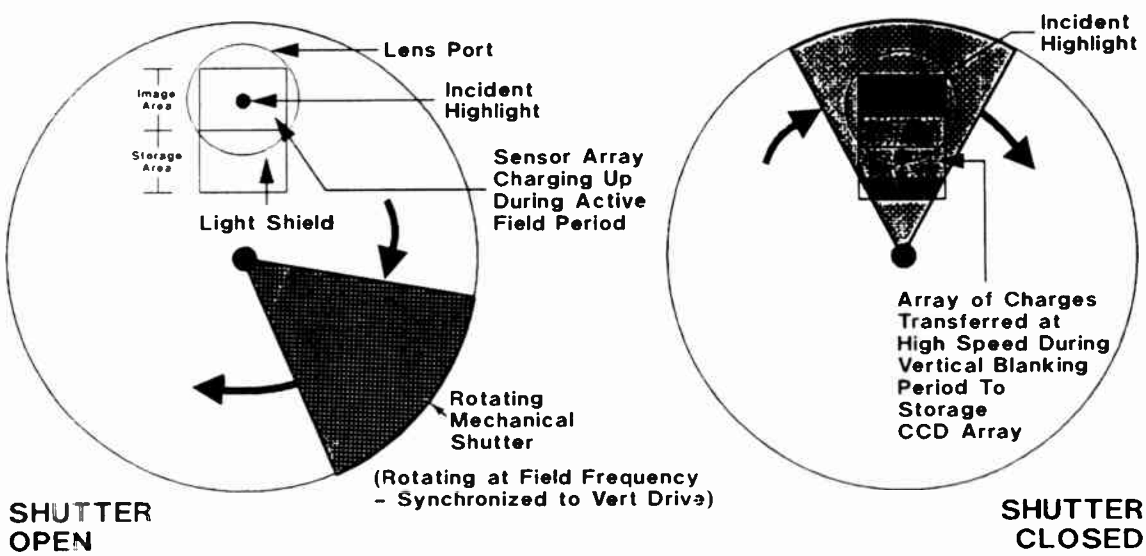


Figure 5.3-6. Removal of vertical smear in the frame-transfer CCD by means of a rotating mechanical shutter.

registers. This reduces sensitivity, which can, however, be recovered by other means. The use of the separate interline storage register now makes it possible to transfer the charge packets out of the sensor array without worrying about incident light directly contaminating these charges; but there is still one more problem to contend with: the intense highlight. When an intense highlight, such as the filament of a car headlight, is focused on the sensing element of the imager, a small amount of this light is reflected within the CCD and will be able to add charges to the immediately adjacent interline storage register. This means the vertical smear phenomenon is still present, albeit at a much lower level than the FT CCD, approximately 105 dB below normal output (see Figure 5.3-8). A vertical streak appears above and below the object or light source. A typical example of where this effect would be observed is when the headlights of an oncoming car are shot with a CCD camera at night. The smear is caused by direct leakage of the incoming light into the vertical storage register (path *A*) or via electrons generated through the photoelectric effect (path *B*) into the vertical storage register. While not perfect, the IT imager achieves the long sought goal of a true solid-state imager with performance suitable for many broadcast applications.

Note that there is currently no industry agreement on a method of measurement for the vertical smear CCD artifact. The method used by some manufacturers uses a square highlight at the center of the screen with an area equal to 10% of the active picture area. Based on the threshold of 2 IRE units as the level below which an artifact is not considered objectionable in the active picture area, a CCD camera with 85 dB smear rejection can be exposed to a highlight about 5-1/2 *f*-stops above 100 IRE units of video before the 2 IRE unit smear level is reached.

Frame Interline Transfer (FIT) CCD Imager

The FIT imager combines the best features of the IT and FT structure and virtually eliminates the vertical smear phenomenon. It is, however, a more complex and hence more expensive semiconductor. Figure 5.3-9 illustrates the basic structure of a FIT type CCD.

In the FIT CCD, the charge packets from the sensor elements are first transferred to the interline storage register, and while still within the vertical interval, these charges are quickly transferred to a frame storage array totally protected from light. Because the charges are present in the interline register only for a very short time, the amount of contamination from an intense highlight is extremely small. In the FIT structure, the vertical smear phenomenon has been suppressed to an insignificant level of 140 dB or more below the main signal level.

Spatial Offsetting

Different from the FT type CCD imager that has almost contiguous sensing elements, the presence of the interline storage register in the IT and FIT type CCDs, reduces the area available to the active sensor element to a fraction of the overall pixel area. A reduced sensing area carries the penalty of reduced sensitivity, but allows the use of *spatial offsetting,* an extremely powerful method to increase the effective luminance resolution of the camera. When spatial offsetting is used, the red and blue CCDs are each bonded to their respective optical ports with extreme precision such that they are physically displaced by one-half pixel distance horizontally from the green CCD (see Figure 5.3-10).

Spatial offsetting does not increase the resolution of the individual red, green and blue channels, but spatial offsetting significantly increases the effective number of samples in the luminance channel and thereby increases the effective resolution of this signal. An example of the benefit derived from using spatial offsetting techniques is as follows:

A camera using three 768 pixel imagers that are co-sited (no spatial offset) achieves a limiting luminance resolution of 560 TV lines. When spatial offsetting is used in a camera using three 768 pixel imagers, a limiting resolution of 700 TV lines is achieved for the luminance (Y) channel. A black triangle on a white

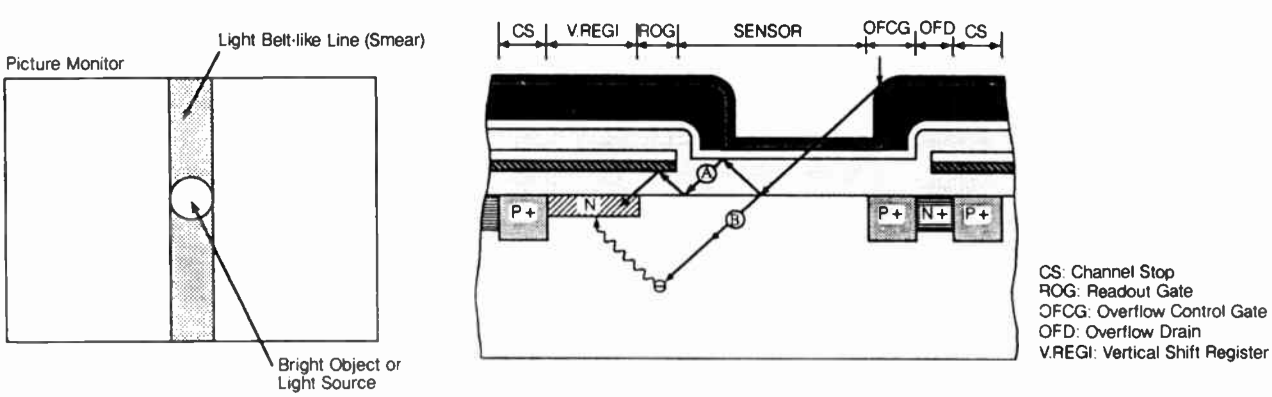


Figure 5.3-8. Vertical smear in an interline transfer CCD. This characteristic of CCD cameras is observed when a very bright object appears in the scene.

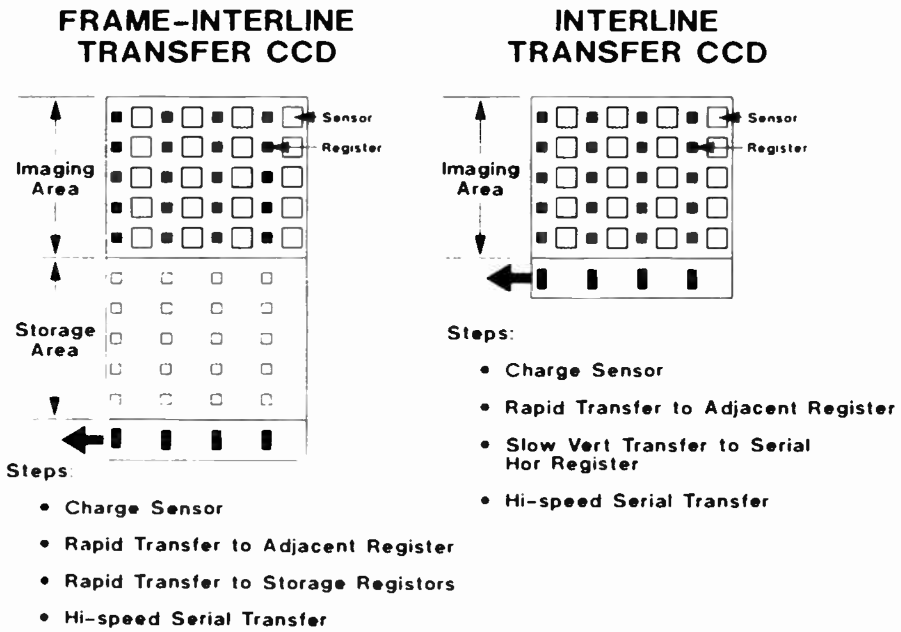


Figure 5.3-9. Steps in the transfer of pixel data for FIT and IT CCDs.

background is used to illustrate the effectiveness of spatial offsetting to enhance the effective resolution of the luminance channel. In Figure 5.3-11, the resulting image projected on the green imager is shown in close-up (a) the corresponding image projected on the red and blue imagers are shown in (b). The quantities of charges corresponding to the input signal stored in the individual sensors of each CCD can be found in the waveforms labeled 1-7. Thus, the output of the green CCD can be visualized as shown in (a) and the outputs of the red and blue CCDs can be visualized as shown in (b). The luminance signal is derived by linear summing of the red, green and blue signals and is shown in illustration (c). The diagonal line in the original scene is reproduced with relatively coarse steps in the individual red, green and blue channels, whereas the diagonal is reproduced with much finer steps in the luminance output, a clear indication of the enhancement in resolution that occurs in the luminance channel when spatial offsetting is used.

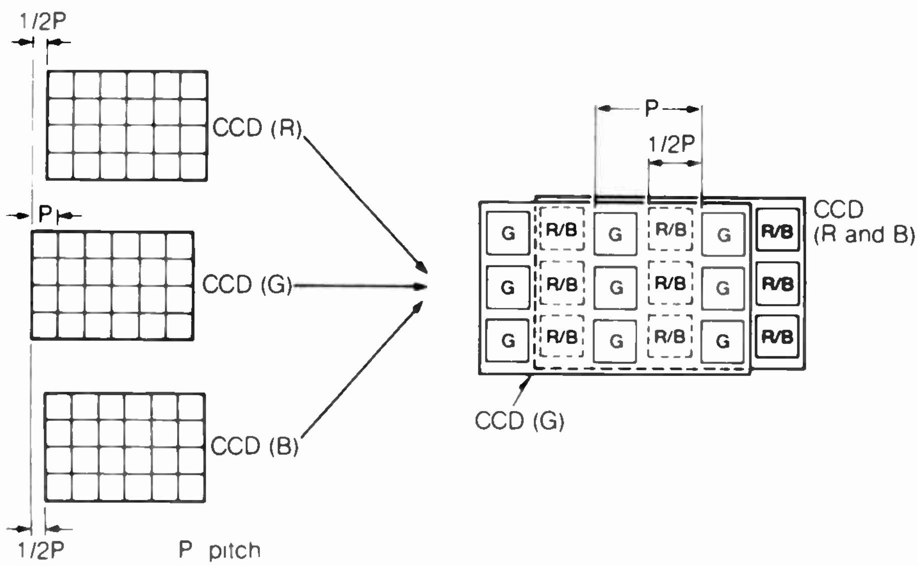


Figure 5.3-10. Spatial offsetting used in CCD cameras.

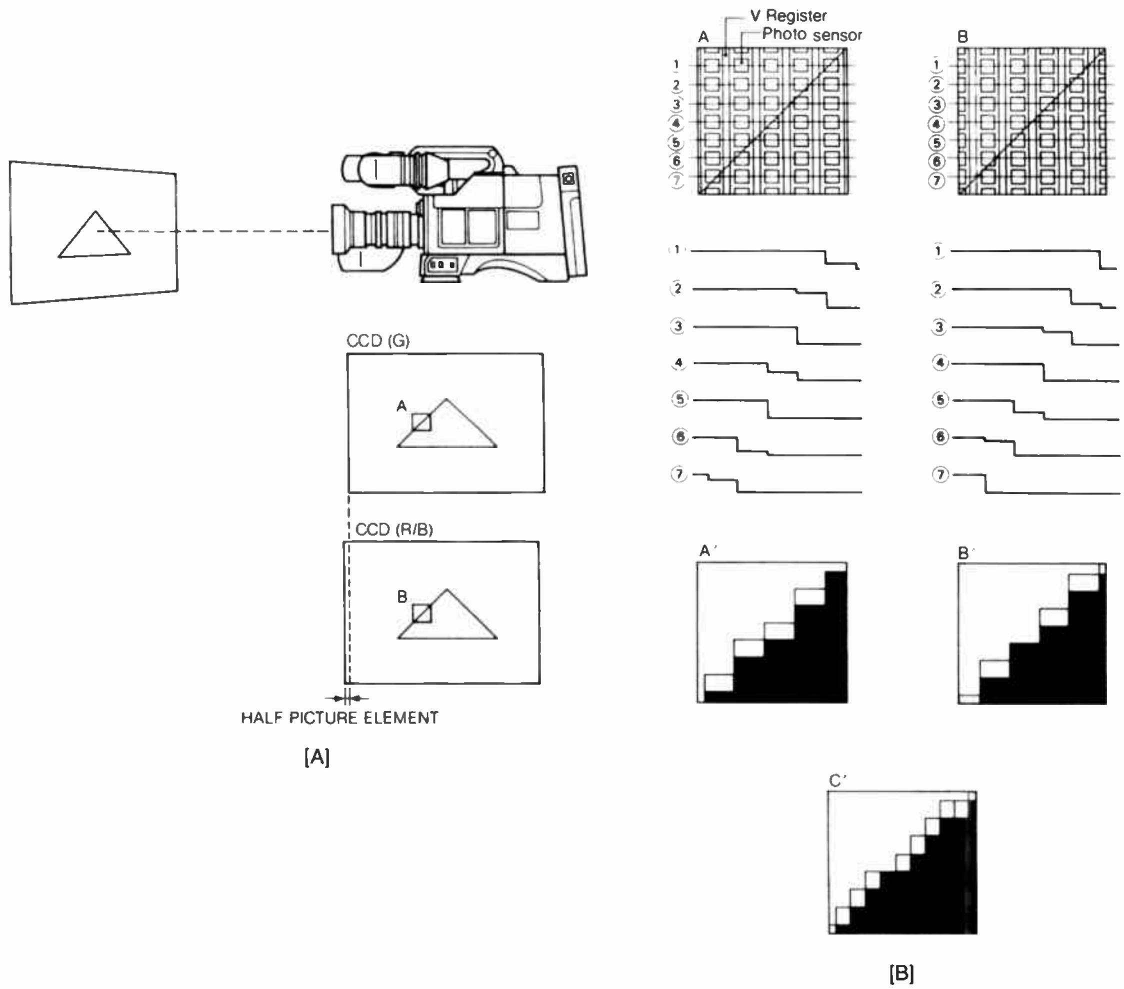


Figure 5.3-11. Enhancement of diagonal resolution by spatial offset of the green image with respect to the red and blue images.

On Chip Lens Array (OCL)

A relatively new word in the vocabulary of CCD terms, the *on chip lens* (OCL) construction, represents a major advance in CCD technology. Sensitivity of 2/3 in. IT and FIT CCD cameras is increased from $f/5.6$ to a remarkable $f/8.0$ at 2000 lux. This represents a threefold improvement in sensitivity compared to a 2/3 in. Plumbicon imager rated at $f/4.5$ and a fourfold increase in sensitivity compared to a camera equipped with Saticons rated $f/4.0$ at 2000 lux.

In an IT and FIT CCD imager, the area assigned to one pixel must be shared by the active sensor, the adjacent interline storage register and the necessary interconnect wiring. As a result a significant part of the incoming light falls on an area of the pixel that does not contribute to accumulation of charge. With the OCL construction, the sensitivity of IT and FIT CCDs is dramatically increased by carefully positioning an array of microscopic lenses, one lens per pixel, over the imager to gather light over a much larger area and focusing it onto the photosensitive area of the pixel (see Figure 5.3-12). By recovering light previously lost, almost one f-stop improvement in sensitivity is achieved. In addition to an increase in sensitivity, the OCL construction also reduces the amount of oblique rays that cause smear in a CCD and thereby dramatically improves the highlight handling capability of the imager. A slight increase in the diffusion of light is the only negative characteristic of the OCL construction and is far outweighed by the positive benefits.

Camera Signal Processing

The output signal from the camera imagers requires extensive processing to produce the video output of a camera. Different from VTRs, where digital recording was implemented long ago to overcome the noise and nonlinearities of the recording tape, analog signal processing has been implemented in cameras to a level of perfection that is to this day the standard of excellence for a digital implementation. Only progress in semiconductor technology and the recent development of

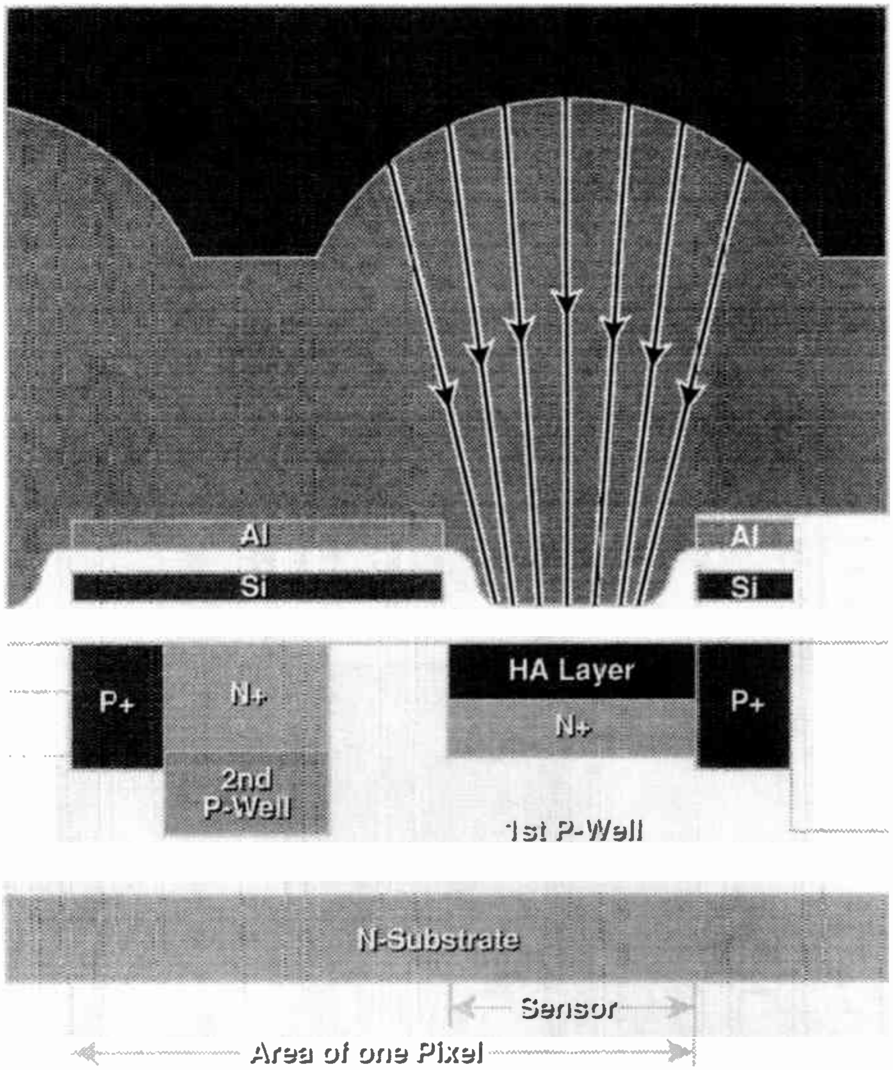


Figure 5.3-12. On chip lens construction. The lens collects light that would otherwise fall on insensitive areas and concentrates it onto the imaging part of the pixel. *(Courtesy of Sony.)*

12 bit analog-to-digital (A/D) converters allows digital signal processing (DSP) cameras to equal and exceed the performance of the best analog signal processing (ASP) cameras. Although it has been difficult to equal and exceed the picture quality of ASP with DSP, DSP provides significant operational benefits. The transition to DSP that started in 1994 will be complete in a few years.

The basic block diagram of an analog signal processing camera is shown in Figure 5.3-13(a). The corresponding block diagram of a digital signal processing camera is shown in Figure 5.3-13(b). Note that the digital processing duplicates the processing steps of the analog processing albeit in the digital domain. This section will elaborate on the analog signal processing followed by some special considerations applicable to a digital signal processing camera.

Analog Signal Processing Camera

The camera processing circuitry has four major functions:

- Adapt to the normal variations in the scene illumination and content
- Correct for unavoidable errors in the lens and optical system
- Correct for the limited dynamic range and other limitations of the video recorder, transmission system and display devices
- Introduce intentional preemphasis, such as gamma correction, to correct for a known nonlinearity in the cathode ray tube.

The camera processing stages implement these specific tasks:

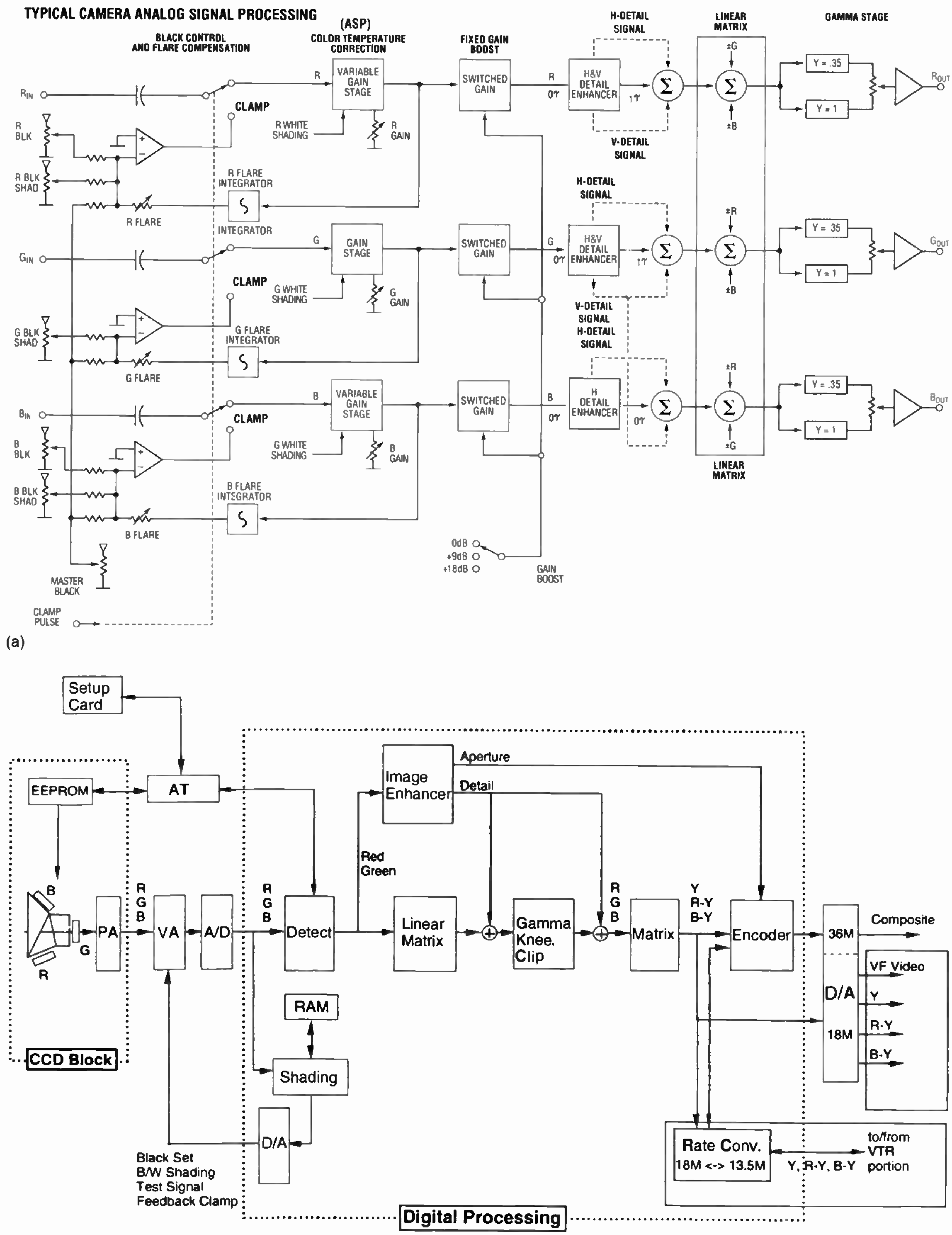


Figure 5.3-13. Camera signal processing system: (a) analog system, (b) digital system.

- **Black level/auto black set circuitry.** The residual signal level from a CCD imager corresponds to the absence of all light or pure black. This *black current*, which is a consequence of the electron mobility (a quantity common to all semiconductor devices), increases with increasing temperature. For this reason, a small section of the CCD imager is masked to protect it from light. The black current reference is determined from masked pixels and used to maintain accurately a constant camera black level with variations in temperature. The overall camera black level can be adjusted using the master black control. The individual red-black level and blue-black level controls can be used for manual black balance or the *auto black* circuit, normally provided in the camera, can be used to set the camera black balance automatically.
- **Shading.** Shading correction circuitry compensates for errors in the lens, color separation optics and imaging devices. Both black and white shading correction controls are normally provided in a professional camera. One example of a shading error is the nonuniform light transmission of a lens. Light transmission of a lens tends to be strongest at the center and falls off towards the edges. White shading (parabolic) can be used to reduce this error significantly.
- **Flare Correction.** The flare correction circuitry provides an approximate correction for the scattering of peripheral rays in the various parts of the optical system. When flare occurs, a small amount of light is scattered into areas of the picture that should be totally devoid of light, resulting in a general rise of the picture black level. The amount of scattered light (rise in black level), tends to increase as the overall amount of light in the scene increases. In the flare correction circuit, an integrator is used to measure the total amount of light in the scene, as an indication of the amount of scattered light. A portion of this integral is then introduced into the camera black level circuits to restore black areas of the picture back to true black (capped level black).
- **Auto White and Color Temperature Correction.** This circuitry is used to correct for the natural color temperature variations of the scene. The red and blue processing channels are equipped with a voltage controlled gain stage that makes it possible to adjust the camera white balance using the red and blue gain controls. An *auto white* circuit, which establishes a white balance automatically, is normally also provided in the camera.
- **Gain Boost.** This circuitry increases the effective sensitivity of the camera, at the expense of creating a noisier picture, by introducing at least three fixed electronic gain steps, typically totaling +9 dB or +18 dB, into the processing circuitry.
- **Linear Matrix.** The primary function of the linear matrix is to correct for deficiencies in the color separation process. The spectral characteristics of the optical system do not provide negative lobes prescribed by the ideal spectral characteristics for the color separation process. A matrix with negative and positive coefficients is used in the signal processing stages to correct for the missing lobes (see Figure 5.3-14).

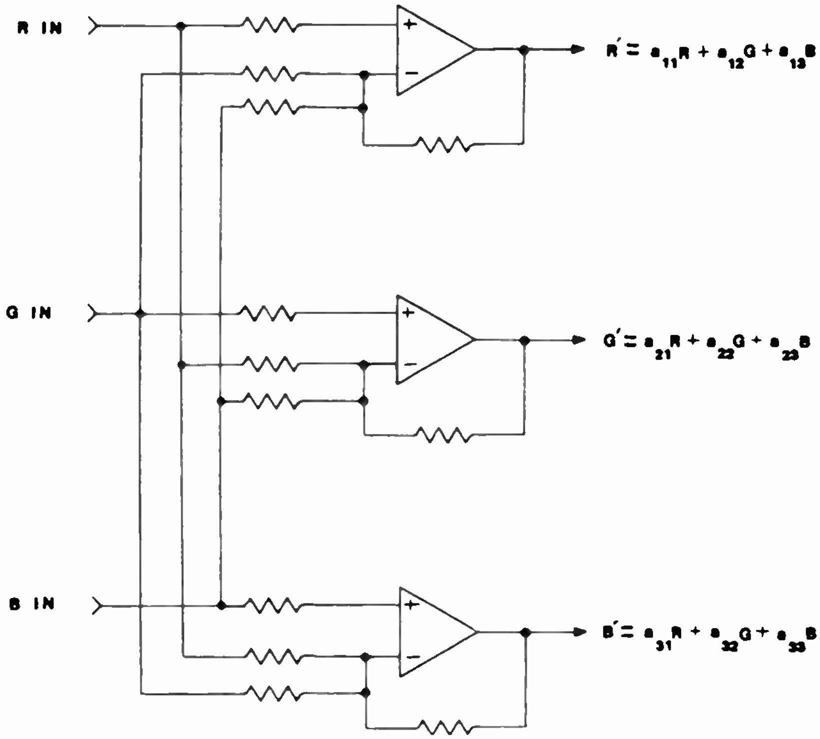


Figure 5.3-14. Basic diagram for the linear matrix circuit.

If an adjustable linear matrix instead of a fixed matrix is provided in the camera, it is possible to match the colorimetry of two dissimilar cameras by judicious adjustment of this matrix.

- **Image Enhancement (Contour Correction or Detail Correction.** This circuitry increases the subjective crispness of the television picture. A detail signal is used to outline transitions in the image, resulting in an apparent increase in the crispness of the picture. A similar outlining of transitions does not occur in film. The closest parallel is in graphic arts, where the apparent crispness of a graphic is increased by outlining the transitions of text or drawings to make them stand out. The dramatic improvement in subjective picture quality achieved by the detail correction circuit is clearly shown by turning the circuit on and off in a working camera. High end cameras generate enhancement for both horizontal and vertical transitions, and this type of image enhancer is called a *two line* enhancer. To provide enhancement for both horizontal and vertical edges, it is necessary to delay the signal from the imager for two full horizontal line periods. Extensive circuitry, including two expensive glass delay lines, is required and hence the designation *two line* enhancer. Less costly cameras frequently provide enhancement for the horizontal transitions only. This type of enhancer is typically called a *one-line enhancer* because it avoids the high cost of the horizontal delay lines.

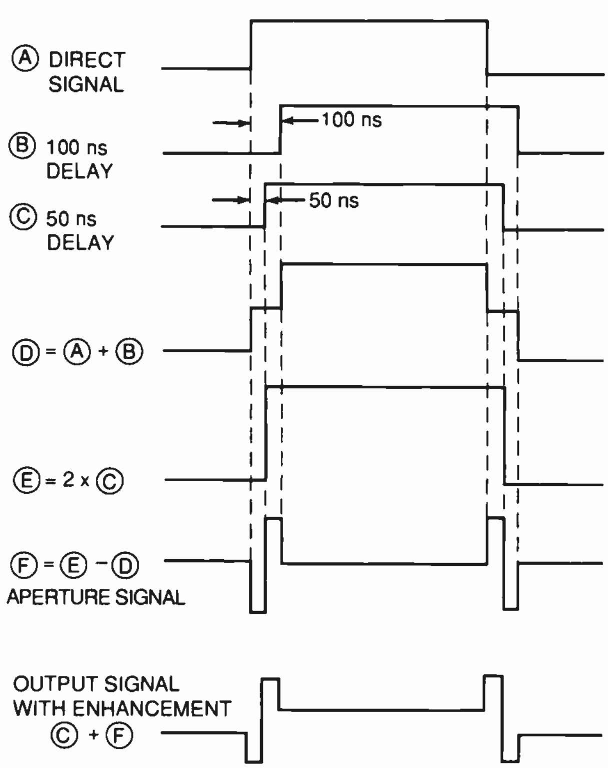


Figure 5.3-15. Generation of the horizontal enhancement signal.

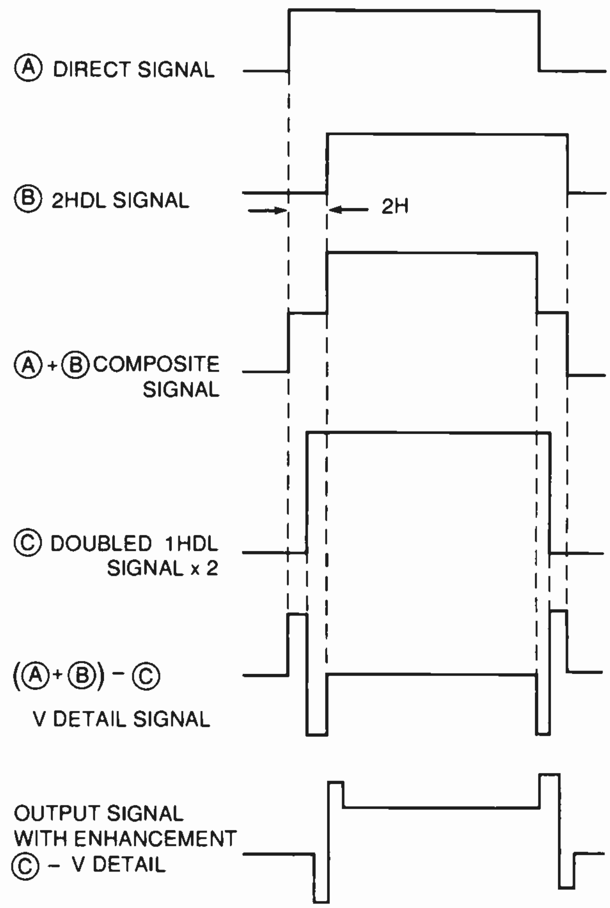


Figure 5.3-16. Generation of the vertical enhancement signal.

The detail circuit generates the detail signal, and it is one of the most complex circuits in the camera (see Figures 5.3-15 and 5.3-16).

- **Gamma Correction.** This circuit precorrects for the nonlinearity of the CRT used in a TV set or monitor by boosting the signals near black and compressing the signal in the white part of the picture, the exact inverse of the nonlinearity of the CRT. The overall result of applying a signal with gamma correction to a CRT is a faithful reproduction of the gray scale in the original scene, as illustrated in Figure 5.3-17.
- **Automatic Knee, Knee and White Clip.** Outdoor scenes routinely encompass contrast ratios as high as 1000:1. One of the most difficult tasks of the video camera is to capture the important content of the scene, while compressing the unimportant details to fit into the limited 40:1 contrast range from camera to the home TV set that is supported by the overall conventional television system.

Significant progress has been made in the ability of the video camera to handle the large dynamic range of outdoor scenes. The hard white clip at 105 IRE implemented in early cameras resulted in a loss of

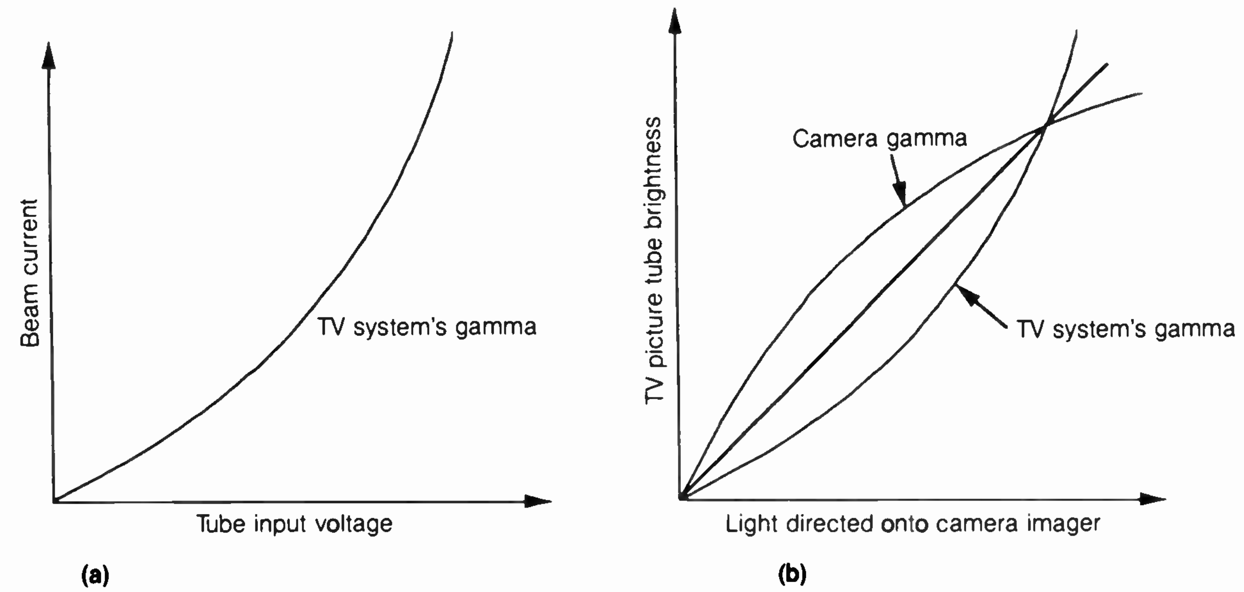


Figure 5.3-17. Gamma correction process: (a) CRT gamma, (b) gamma correction.

significant picture information in the highlight areas. The next advance in handling was the *knee circuit:* a soft bend or knee introduced, while the hard knee was moved up to about 108 IRE. At least one additional *f*-stop of highlight detail was compressed in the range from 96 to 108 IRE. The most recent advance in camera highlight handling is the *automatic knee* circuit, also called *dynamic contrast control* (DCC) or *dynamic knee* (see Figure 5.3-18). With this circuit, the onset of the soft knee is progressively shifted as low as 85 IRE with the existence of strong highlights. The equivalent of up to 600 IRE of picture information in the highlight area is now compressed in the range from 85 to 108 IRE of video (the range of video that can be recorded by a VTR, transmitted and reproduced by a display device).

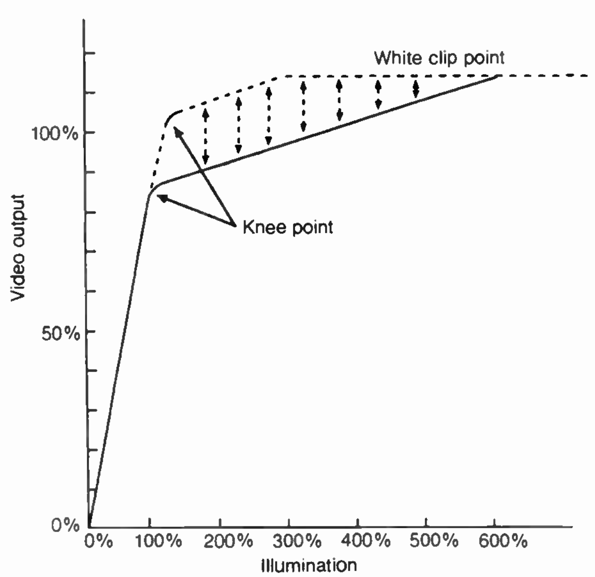


Figure 5.3-18. Auto-knee dynamic contrast control.

Auto Iris and Zebra Circuits. In a production environment, significant time and effort are expended to adjust the lighting and iris opening for the desired artistic effect. Manual adjustment of the iris with the support of a waveform and picture monitor is most commonly used in this application. The *auto iris* circuit in combination with the *zebra* indication are the tools provided by the camera for ENG and field production, where it is not practical to use a waveform monitor and picture monitor. The basic auto iris circuit evaluates the video level, compares this level to a reference (typically called the *auto iris set point*), and then sends a correction signal to the lens to open or close the iris as appropriate to achieve proper exposure. Various designs and degrees of sophistication are used to improve the performance of the auto iris circuit in difficult lighting situations. The method used to evaluate the video level is the key to improved performance. Some of the improvements to the basic auto iris circuit are:

- **Peak/average picture evaluation.** The peak level of the video signal is detected by one circuit, the average video level is detected by a second circuit, and a mix of the two circuits (peak/average adjustment) is then used for improved iris control.
- **Weighting and masking.** Commonly used to improve the performance of an auto iris circuit. Because most of the important action takes place at the central part of the picture, correct exposure at the center is most important. Without weighting or masking, a bright light source at the periphery of the picture, such as the sky, may cause the iris to close and provide improper exposure at the center. When

bright highlights at the edges of the picture are excluded from the picture evaluation by masking, or the video level at the center of the picture is weighted more heavily than video at the periphery, performance of the auto iris circuit with difficult scene content is significantly improved.

Despite the various refinements in picture evaluation, the resulting auto iris setting is not satisfactory with some difficult scenes. The zebra circuit can then be of significant help in finding the proper manual iris setting by superimposing diagonal stripes on bright areas in the viewfinder image. There are two common operating modes for the camera zebra circuit. In one mode, it is adjusted to identify in the viewfinder all video at the 60 to 70 IRE level (proper exposure for facial features). The other operating mode of the zebra circuit is to identify all video that is at 100 IRE and above (all video that is going into the white clippers). In the latter mode, the iris is adjusted until the brightest part of the scene that still carries useful information is just below the threshold of the characteristic diagonal zebra stripes.

- **Encoder.** For some applications of a video camera, the *component* video signals in the form of R/G/B or Y/R − Y/B − Y are used directly. For other applications, it is desirable to encode the red, green and blue signals generated by the camera into a *composite* color signal. In the United States the camera signals are normally encoded to the NTSC standard.

The encoder used in a high performance video camera differs from other commonly used encoders in having exceptionally wideband luminance response, reaching as high as 10 MHz. Figure 5.3-19 shows a simplified block diagram of an encoder.

The RGB output of the camera processing system is applied to the luminance or Y matrix and the I and Q color matrixes. The equations of these three matrixes are:

$$Y = 0.30R + 0.59G + 0.11B$$

$$I = 0.60R - 0.28G - 0.32B$$

$$Q = 0.21R - 0.52G + 0.31B$$

The I and Q signals are processed by 1.2 Mhz and 0.6 MHz lowpass filters respectively. The Q filter introduces a delay of about 750 ns, which introduces delay into both signals. The I signal filtering is less severe and

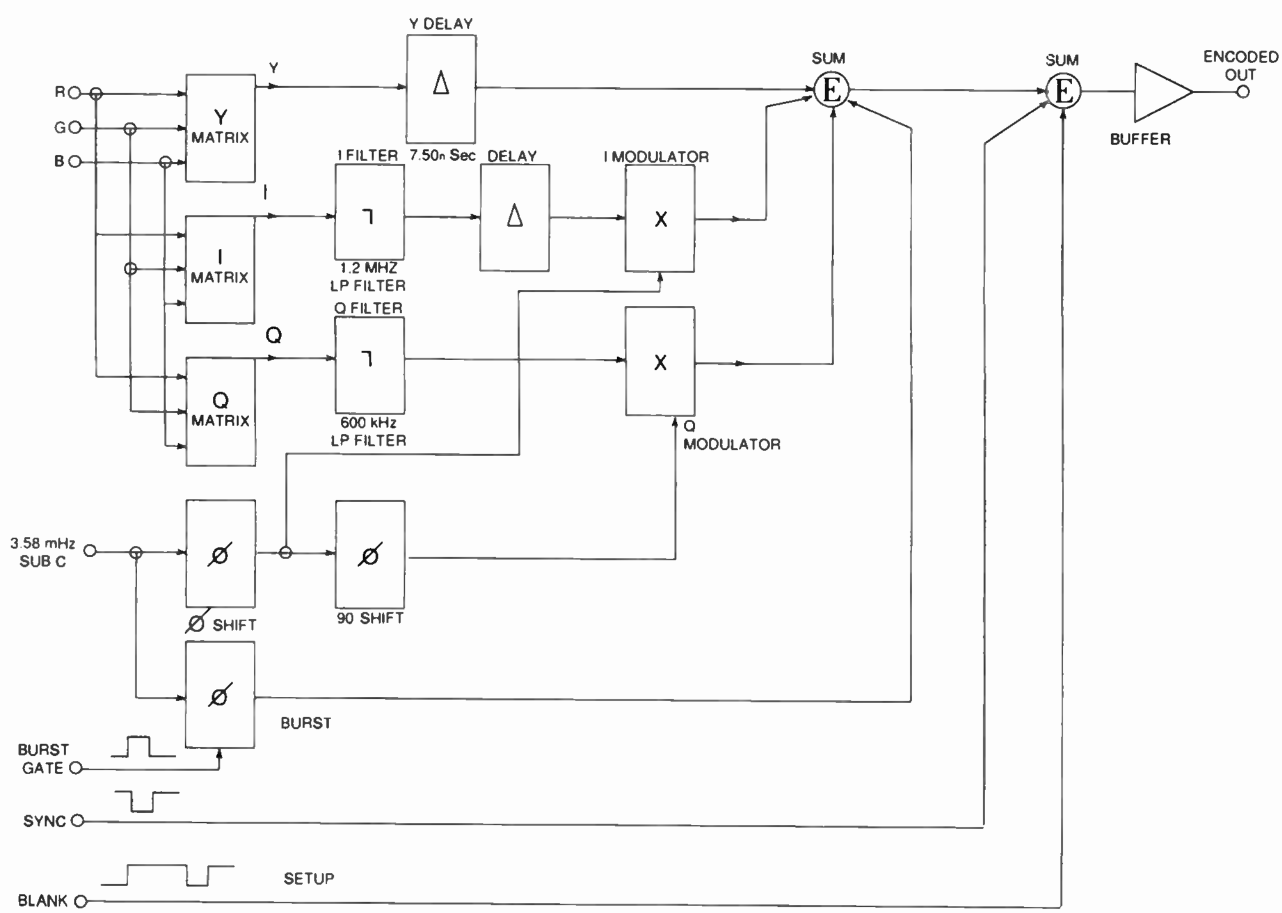


Figure 5.3-19. Simplified block diagram of an NTSC encoder.

has correspondingly less delay. An additional delay or a special filter design is used to equalize the delay of the I and Q channels. The bandwidth-limited I and Q signals are then applied to the I and Q modulators, where the I and Q information modulates two 3.58 MHz subcarriers in quadrature (phase shifted by 90°). The output of the Y matrix is delayed by about 750 ns to match the inherent delay in the I and Q chroma channels. The output of the I and Q modulators, the 3.58 MHz burst, sync and setup are added to the Y signal to create the composite NTSC output.

Normally, a color bar generator is included in the encoder to aid in system setup. The typical output waveform of an NTSC encoder is shown in Figure 5.3-20 (for a color bar signal). The phase relationship of the I and Q axis vs. burst is shown in Figure 5.3-21.

- **Sync/Timing Generator.** This is the master timing system that provides the various timing waveforms required by the camera, in accordance with the television standard used. In the United States broadcast cameras are normally designed so that the output signal meets the requirements of the Electronic Industries Association recommended standard RS-170M.

The timing waveforms required by the camera circuits are normally derived by digital countdown from a crystal oscillator operating at four times subcarrier frequency ($4f_{sc}$) or 14.3 MHz. A temperature-compensated crystal oscillator is used to maintain the strict frequency requirements of the RS-170M Standard when the camera is used in a self contained mode. For multicamera operation, it is necessary to synchronize the cameras. For this purpose, the camera sync/timing generator is provided with genlock capability. The sync/timing generator has the ability to phase lock to an external composite video signal applied to the camera genlock input.

- **Power Supply.** The power supply voltages required to operate a solid-state camera are quite low, and

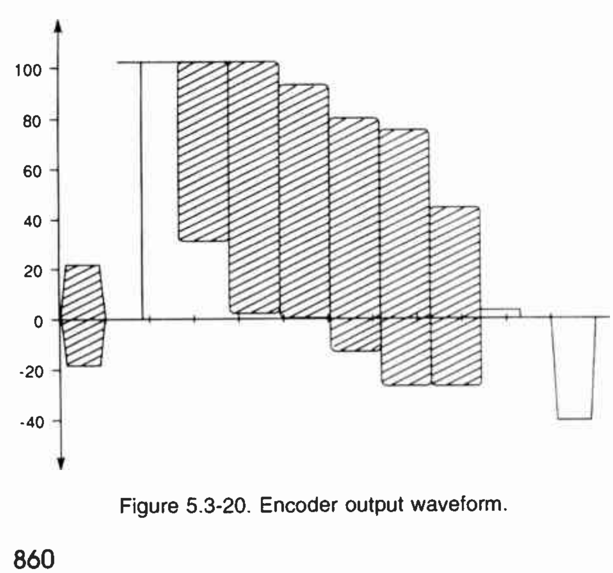


Figure 5.3-20. Encoder output waveform.

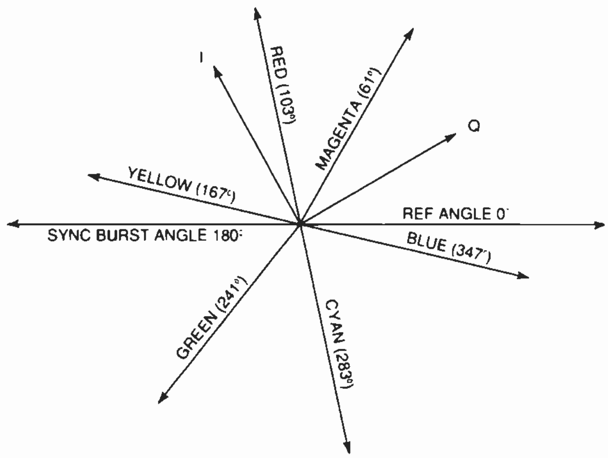


Figure 5.3-21. Subcarrier phase relationship in the NTSC encoder.

only normal precautions need to be observed when working on the inside of a CCD camera. The accelerating voltages used in the camera viewfinder are the exception, of course. Tube cameras, on the other hand, require substantial accelerating voltages, and appropriate precautions should be observed when working on any viewfinder or a tube type camera.

Virtually all modern cameras use a switching type regulator that converts the incoming voltage to the various potentials required by the camera. The universally accepted power supply range for portable cameras is 10.8 to 17.4 Vdc. By the nature of a switching type regulator, the camera presents a constant power load to a battery or an external power source.

Digital Signal Processing Camera

Digital video processing was first introduced into videotape recording, where the benefits were most dramatic and cost effective. In most portable applications digital processing was either too expensive for the benefits it provided. As a result, a mix of digital and analog equipment represented the most cost effective system configuration. With time, the performance of digital circuits has improved dramatically, while costs have been reduced. It is increasingly cost effective to use digital circuitry in virtually all production equipment, including cameras.

The standard imager for cameras is currently the CCD, a thoroughly analog device despite the fact that the information from the imager is read out in discrete packets. The dynamic range of a CCD (or a pickup tube) is quite large. It is not uncommon for the early processing stages of an analog camera to faithfully process signals as high as 600% video. Digital processing with an inordinate high number of bits per sample (greater than 12 bit A/D) is required to handle this large dynamic range while retaining the ability to resolve fine shades of luminance difference. In addition, although the camera is required to process highlights up to 600% of video, it is only necessary to compress these large highlights to a more reasonable range. An analog pre-knee circuit has been found to execute this function quite efficiently so that it is possible to limit the digital signal processing circuitry to a dynamic range of 226%. Until the current analog imagers are displaced by true digital imagers, studio and portable cameras will remain hybrid analog/digital devices. Figure 5.3-22 illustrates the analog pre-knee with 10 bit linear A/D bit assignment.

The benefits of digital signal processing for video cameras include the following:

- **High stability and reliability.** However careful its design, the circuitry in analog cameras is inevitably subject to drifting that requires manual correction. Subsequent readjustments are then subject to operator interpretation, and so the actual setup of a camera at any one time is difficult to define. With digital processing, parameters are held in a digital memory and stay constant for long periods of time (potentiometers, the least reliable component of cameras, are reduced from about 150 for analog processing to less than 6 for cameras with digital processing). As a result, the need for operator adjustment is dramatically reduced. A further advantage of digital processing is that it is much easier to implement digital circuits in ICs and LSIs than analog circuits, allowing the size and weight of cameras to be reduced.
- **Precise alignment.** The accuracy of a camera setup can be defined with great precision by digital processing. Moreover, variations between cameras, which are difficult to avoid with analog processing cameras, can be reduced to a minimum with digital processing by simply choosing the same setup parameters.
- **Flexible signal processing and parameter setting.** A significant advantage of digital signal processing is that it can provide very flexible operation. Many camera parameters can be controlled and each setting can be controlled over a wide range of values. It has been well known that different camera adjustments can dramatically improve the ability of the device to capture difficult scene material. With analog cameras, such custom adjustment is difficult and time consuming to implement as well as to restore. With digital processing cameras, control adjustments can be manipulated quickly and easily, stored and then recalled with great accuracy. Convenient menus are provided to manipulate the digital parameters and some cameras are now equipped with miniature memory cards that allow preset camera adjustments to be executed almost instantaneously by the simple expedient of reading the setup card.
- **Improved reliability.** One of the most important concerns of the broadcaster is the reliability of equipment. This is considerably improved with digital processing. In digital cameras, all the processing parameters can be completely and precisely defined at the design stage and almost all the potentiometers of analog cameras are replaced by semiconductor memories. Consequently, the digital cameras show precise and stable performance regardless of environmental conditions such as temperature, humidity, shock and vibration.

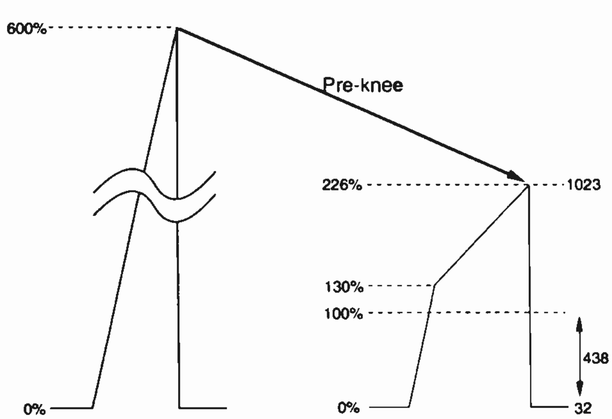


Figure 5.3-22. Pre-knee and 10 bit linear A/D bit assignment.

Interlaced and Progressive Scan Cameras

Starting in the mid 1970s, solid-state CCD imagers began to take the place of pickup tubes in video cameras. And more recently, advances in high capacity, low cost memory storage, have made it possible for cameras generate various types of output signals to suit different applications.

The most promising of these new camera output capabilities is *progressive scan.* Progressive scan images capture amazing clarity, without the motion double exposure, or flicker that is seen in still-frame images derived from conventional cameras that use interlaced scanning. The clear, flicker-free image quality provided by progressive scan cameras offers dramatic benefits for anyone who wants to view a still frame of a moving scene, use the captured video as the base line signal for extensive manipulation by a computer, analyze details of the image or print out a hard copy of the video frame.

To understand the creation of interlaced scanning it is best to look in time at the operation of the pickup tube and the cathode ray picture tube.

In the video camera, an electron beam is used to scan the visual image. In the television monitor, the picture tube uses an electron beam to scan the phosphor layer deposited on the inside of the front surface, thereby creating a faithful reproduction of the image captured by the camera.

To create the television image, the electron beam starts at the upper left-hand comer of the pickup tube scanning pattern, or raster, and moves from left to right. When the scan reaches the right-hand edge of the picture, the beam retraces quickly back to the left edge and drops down to the next scan line in the raster. This process is repeated until the complete picture has been scanned.

The picture information that results from this scanning process is transmitted to the TV monitor, where an electron beam in the picture tube is synchronized to trace out the same scan pattern that was used to scan the original image in the pickup tube. When the electron beam hits the phosphors on the front of the picture tube, a light output proportional to the intensity of the beam is generated.

Engineers quickly discovered that, if the pickup tube and the CRT are scanned progressively, line by line, the light output of the phosphors at the top of the picture tube had already started to fade by the time the scanning beam reached the phosphors on the bottom of the tube. As a result, progressive scanning can exhibit objectionable flicker unless the number of complete pictures scanned each second (*refresh rate*) is quite high.

Increasing the refresh rate overcomes the flicker problem and improves the ability of the system to portray motion more accurately, but a higher number of complete scans, or frames, requires a higher transmission bandwidth.

To reduce flicker and reproduce smooth motion, and at the same time minimize the amount of bandwidth required, the creators of television developed a strategy called *interlaced scanning*. In the NTSC interlaced scanning system, the TV camera takes 60 independent pictures (or samples) of the scene per second—but only half of the full 525 scan lines are transmitted for each picture that is captured. This nominal 262.5 line sample is called a *field*, and contains half of the original picture information. The two successive fields are respectively called the *odd field* and the *even field*. When combined in the CRT, the two fields create one complete picture or *frame*.

Interlaced scanning eliminates flicker and reduces the amount of transmission bandwidth required to achieve smooth motion compared to progressive scanning, but the ability to extract a clear, flicker-free still-frame from the video signal is significantly impaired compared to a video system that uses progressive scanning at the same frame rate.

The new digital transmission system adopted by the FCC allows both interlaced and progressively scanned images to be transmitted.

The High-Definition Camera

The performance of the current NTSC television system using 525 scan lines is limited. Although the new digital transmission standard that conveys three separate color signals will improve the picture quality to the home, a dramatic improvement in television quality is observed when the number of scan lines is increased to a thousand scan lines and more. Although the technical requirements to produce such a high definition camera are quite severe, the actual design of the HD camera is essentially identical to the standard-definition camera described previously.

BASIC CAMERA PERFORMANCE VERIFICATION

With tube type cameras, it was almost mandatory to fine tune the camera before a major shoot to achieve optimum performance. The intrinsic stability of the CCD imager and the stability of the circuitry used in current CCD cameras now make it possible to operate the camera for several months without internal readjustment. Physical damage to the camera or lens, in use or transport, is probably the most frequent cause for a loss of performance in a current CCD camera. With careful handling, the probability of malfunction is very small. It is nevertheless prudent to schedule a quick checkout of the camera before the start of a major shoot, when the high cost of talent and other aspects of the production are considered.

The following items are appropriate for inclusion in such a quick checkout procedure. If the test results show a significant deviation from the manufacturers specifications or from the data previously obtained for the same test, a more thorough examination of the camera, as prescribed in the camera service manual, is then indicated. Several basic tests with expected data are also described. Adjustments appropriate to a CCD cameras are also described.

Visual Inspection And Mechanical Check

Visually inspect the camera and lens for any evidence of physical damage as a clue to the possibility of more serious internal damage. Carefully operate the lens adjustments—manual and servo zoom, focus and manual iris. If there is evidence of binding or a rough spot in any of these adjustments, physical damage that may affect the optical performance of the lens must be suspected. Inspect the front and rear lens elements, clean if necessary using pure alcohol and soft, lint-free wipes. Fingerprints in particular should be removed as quickly as possible to avoid harm to the optical coating of the lens elements. Note that the lens manufacturers generally discourage the use of silicon-impregnated wipes for cleaning high quality optics.

Confirmation Of The Camera Encoder

A properly adjusted encoder is particularly useful since it provides a convenient window to look inside the camera and confirm proper operation of the remaining circuitry. Encoder setup is particularly easy to confirm because the color bar generator, normally provided in a professional camera, provides a convenient self test of the encoder. To confirm proper operation of the camera encoder:

- Apply the camera (encoder) output signal to a waveform monitor (WFM), vectorscope and picture monitor (a high resolution black and white monitor with 800 television lines (TVL) or higher resolution is recommended)
- Terminate the WFM, vectorscope and picture monitor using a discrete 75 Ω termination. The preferred tolerance for this termination is ±0.1% and no greater than ±1%. Internal terminations should not be used unless they have been tested and, if necessary replaced with terminations within the recommended tolerance
- Select the color bar mode on the camera. Confirm on the vectorscope that the burst and I and Q vectors

are of the correct phase and amplitude. Confirm that all of the color bar vectors fall within the tolerance boxes on the vector scope.

If all of the above vectors are within tolerance, correct operation and adjustment of the encoder is confirmed.

Confirmation of Auto Black

To confirm proper operation:

- Activate the auto black circuitry of the camera
- Confirm that the lens caps during this adjustment. Confirm the character display in the viewfinder indicates the auto black adjustment has been successfully executed
- Select the 0 dB, +9 dB, and +18 dB gain settings in sequence and confirm the black level adjustment is correct for all three gain settings. This is most easily confirmed with the vectorscope; the output signal should be a dot at the center of the display with no shift in position as gain is switched. The only change should be an increase in noise at the higher gain settings.

Lens Back-Focus

This adjustment trims the lens to the specific optical dimensions of the camera. Whenever a new lens is put on a camera, it is necessary to make this adjustment. Some lenses use a screwdriver lock, while others use a knurled knob to lock the lens back-focus adjustment in place. Accidental misadjustment in use has been known to occur, and it is therefore recommended to confirm this adjustment. To adjust the lens back-focus:

- Locate a Siemens Star Chart (Figure 5.3-23) at least 10 ft from the camera. Place the chart in a location with low lighting such that the lens iris is wide open
- Using the high resolution picture monitor: 1) adjust the lens zoom for full close-up and adjust for best focus using the focus ring on the front of the lens, 2) and adjust the zoom for maximum wide angle position and adjust for optimum focus using the lens back-focus adjustment. Repeat both steps several times
- Securely lock the lens back-focus adjustment in place
- Confirm the lens stays in focus over the full zoom range.

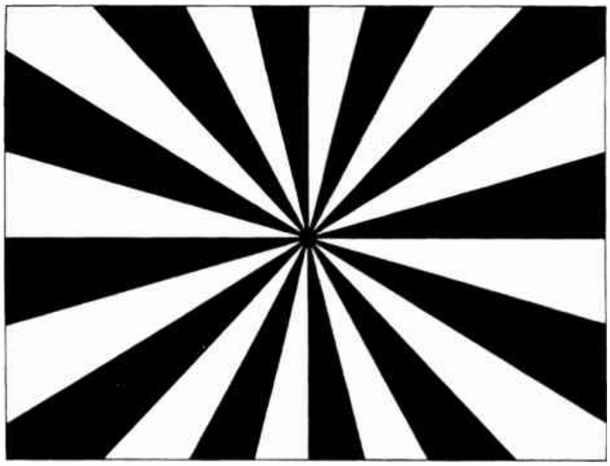


Figure 5.3-23. Siemans star chart.

Black Shading

To confirm black shading:

- Cap the lens. Raise the master black level to about 10–12 IRE to avoid any clipping
- Using the waveform monitor in the vertical display mode, then in the horizontal display mode, if the black level is a thin horizontal line, there is no black shading in any of the three color components
- Restore the black level to its proper position.

Detail Circuit

Using the 11-step gray scale chart, confirm the amplitude of the detail signal as required for the application. A stronger detail signal is typically required for a lower performance recorder and less for a higher performance recorder.

Optional Tests

If the camera system is capable of resolving fine detail close to the limiting resolution specified by the manufacturer, there is a strong assurance that the lens, camera optics and overall camera signal processing are working correctly.

Use a suitable chart with resolution wedges or a chart with a multiburst pattern and confirm that the overall resolution of the camera system is close to the manufacturers specification (see Figures 5.3-24 and 5.3-25).

White Shading

To confirm white shading:

- Set up a uniformly lit white test chart
- Using the waveform monitor, open the lens to obtain about 70 IRE units of video (confirm the iris is in the range of $f/4.0$ to $f/5.6$, adjust the lighting if necessary), confirm there is a minimum of vertical, then horizontal, shading
- Adjust as necessary using the camera horizontal and vertical white shading controls.

Flare

The camera flare correction circuitry provides an approximate correction for flare or scattering of peripheral rays in the various parts of the optical system. To confirm the adjustment of the flare correction circuit:

- Frame an 11-step gray scale chart that includes a very low reflectance strip of black velvet added to the chart
- Adjust the iris from fully closed until the white chip is at 100 IRE units of video. The flare compensation circuitry is adjusted correctly if there is almost no

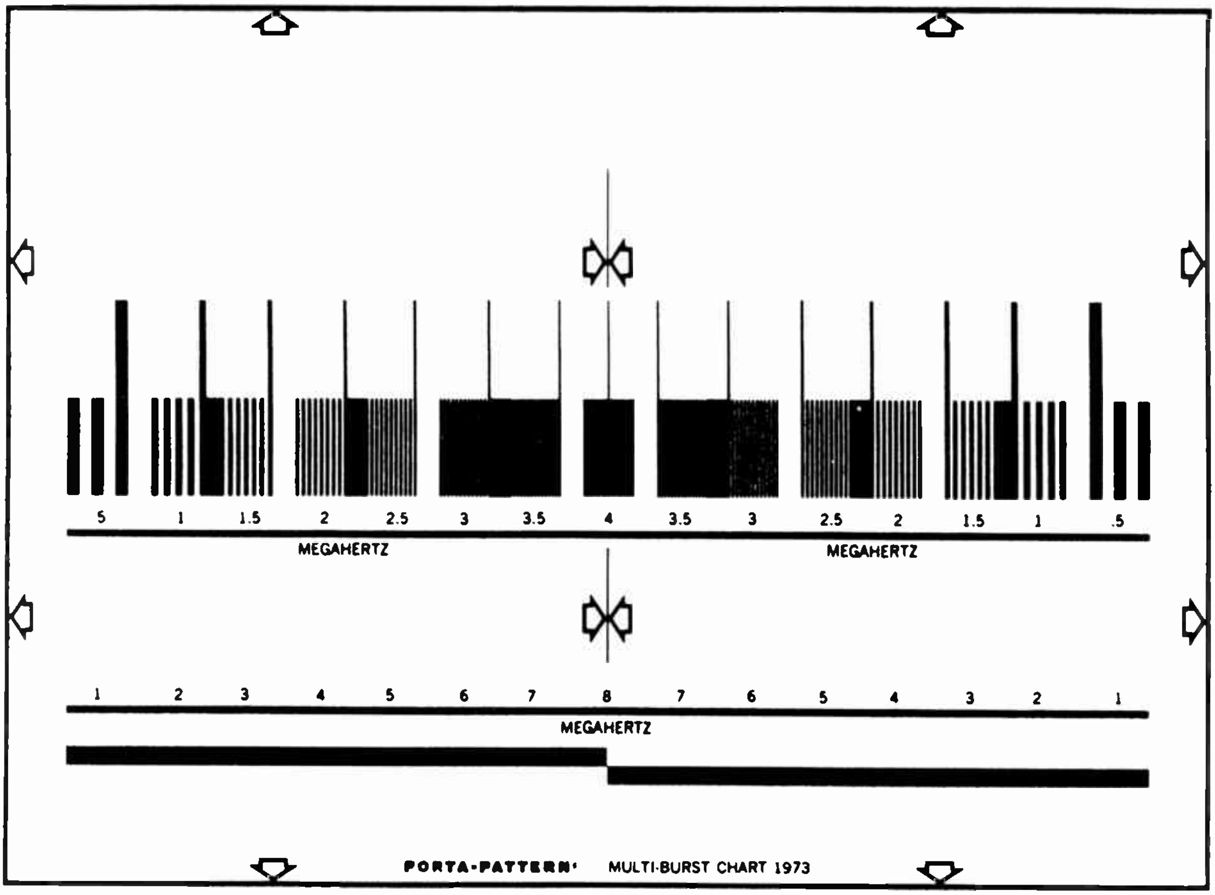


Figure 5.3-24. Multiburst chart (10 MHz).

rise in the black level of the velvet strip as the white level is increased to 100 IRE units and only a small rise in the black level, with no change in hue, when the iris is opened one more f-stop beyond the 100 IRE units point

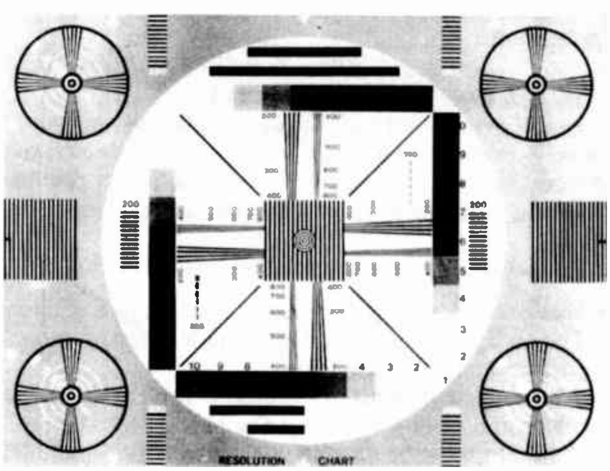


Figure 5.3-25. Resolution wedge chart.

- Adjust the R, G and B flare controls as defined in the camera service manual if the flare correction is not adjusted correctly.

Linear Matrix

When it is necessary to use two dissimilar cameras models in a multicamera shoot, and either of the two models provides an adjustable linear matrix, it is possible to use the variable matrix to obtain a better colorimetry match between cameras. Specific matrix parameters and adjustments (if any) will be found in the camera service manuals.

CAMERA SELECTION

This section is intended to give some practical guidelines as to the types of camera to consider for a given application.

Studio/Remote Applications

Triax cameras have now almost completely replaced cameras using multiconductor cable for studio and field acquisition. The small size, ruggedness and low cost of triax cable, plus the ease of repair, provide a strong advantage over cameras using multiconductor cable. As analog cameras are being replaced by full digital cameras, the bandwidth provided by triax cable is no longer adequate and triax cable will be slowly replaced by camera systems using fiber optic cables.

A *studio camera* system is the system of choice when:

- There can be no compromise in picture quality. The ability to use large studio lenses with superior optical performance is a key factor
- A teleprompter is required for coaching the talent and both power and the signals for the teleprompter must be carried in the camera triax cable
- Extended zoom capability (well above the 18× zoom ratio normally provided and the 32× zoom ratio provided by the largest portable lenses) is required, which can only be achieved with studio lenses
- The look and feel of a large camera head is desired

A *portable camera* is indicated when:

- Mobility and handheld operation is important
- The *reach* (zoom ratio and maximum focal length) of available ENG style lenses is adequate
- The lighter weight of the portable camera head is critical, such as with robotic pedestals
- Teleprompter and multiple channels of intercom are not mandatory, or use of additional cables for teleprompter use is not objectionable.

Electronic Field Production (EFP)

Field production has seen a gradual change from the use of a self-contained camera and a separate VTR to the use of a portable camera that is capable of accepting a dockable VTR creating a two-piece camcorder. With the camcorder, creative freedom is enhanced significantly by the ability to shoot on a tripod or shoulder, as desired. The dockable camera can also be configured with a suitable adapter and operated with a camera control unit (CCU) for convenient remote operation of the camera system.

The one-piece camcorder, that integrates the camera and recorder in one unit, is now becoming the camcorder of choice for field production use due to its outstanding ease of use and high performance. Frequently used in conjunction with a handheld remote control unit, the one-piece camcorder now combines convenience with artistic control.

Electronic News Gathering (ENG)

For the ENG camera operator, weight, size and ease of use are the critical issues, particularly with the trend to reduce ENG crew size. The one-piece camcorder optimizes size, weight and balance to provide the most ergonomic camera for ENG use. For these reasons, the one-piece camcorder is now the virtual standard camera for ENG use.

Camera Specifications

A TV camera performs the complex task of creating an electronic image of a real scene, ranging from scenes with extreme highlights—scenes with large dynamic range that must be compressed to fit within the capability of the TV system—to scenes with marginal illumination. Defining the performance of a camera in a complete but concise set of specifications may be impossible, but camera manufacturers need to provide more complete specifications than are currently provided. It is not unusual to find a low cost camera with virtually the same published specifications as a camera costing significantly more. Actual day-to-day performance, on the other hand, will probably show the more expensive camera to be far superior in handling difficult lighting situations. For this reason the published camera specifications are no more than a basic guide for which cameras to consider for actual evaluation.

It is no longer necessary to limit the choice of camera to the one with the best picture quality since virtually all current professional CCD cameras make high quality images. Such factors as ease of use, cost and operational features can now frequently be the deciding factors in choosing one camera over another for a specific application.

In an actual camera evaluation, it is typically not necessary to spend a great deal of effort to confirm the specifications provided by the manufacturer. Instead it is recommended to expose the camera or cameras considered for purchase to the most difficult shooting situations that they are likely to be exposed to in the intended application(s), and then make the final choice based on the overall advantages of one camera over the others.

APPENDIX A: COLOR SEPARATION SYSTEMS

A color camera contains an optical system that separates the incoming light into its three component colors. Currently, there are three types of color cameras in descending order of performance:

- The *three imager* design with prism type color separation
- The *two imager* design with a beam splitter and color stripe color separation for the chrominance imager
- The *single imager* design using alternate optical filters over the individual pixels of the imager to achieve color separation.

Dichroic mirrors have also been used for color separation in some portable and studio cameras.[2] The dichroic mirror system is not used in current cameras because there is significant loss of light as compared with the prism system.

The prism type color separation system has become the dominant system for use in broadcast cameras and will be described in more detail.

Color Separation System for Three Imager Color Cameras

In front of the prism, several color correction and neutral density filters are mounted on a filter wheel to

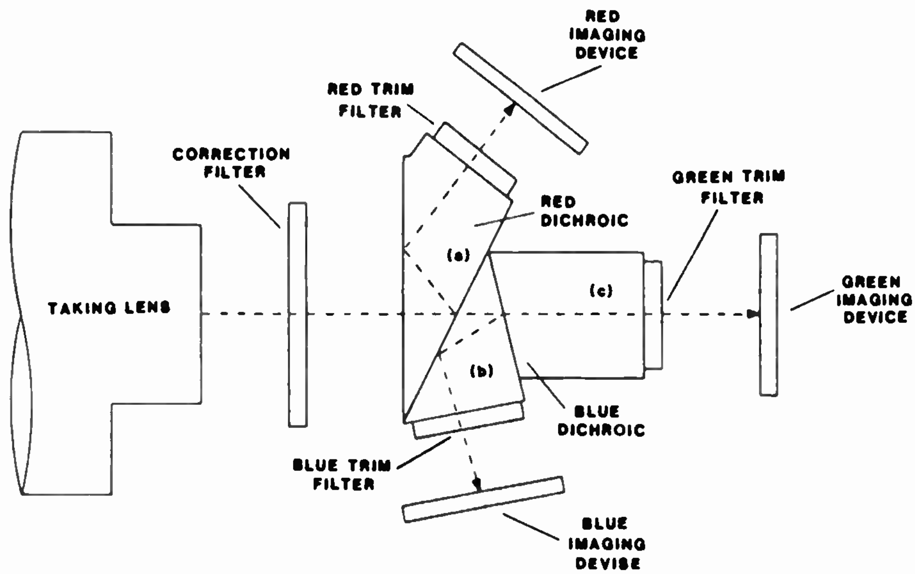


Figure 5.3-A-1. The three imager prism-type color separation system.

provide large-scale compensation levels for the color temperature and intensity of the incoming light. The infrared filter, quarter wavelength filter and anti-aliasing filter described earlier are also located here. The beam-splitting prism system makes use of selective reflection at dichroic layer surfaces to separate the incoming light into the red, green and blue color components. Total reflection, at untreated surfaces, is also used within the prism system (see Figure 5.3-A-1.

A dichroic layer is formed by vacuum evaporation of 10 to 20 layers with alternating high and low refractive indices. Proper choice of the material and thicknesses can give the dichroic layer the property of reflecting only one color and passing other colors.

The spectral characteristics of the prism are the determining factor in the quality of color reproduction achieved by the camera. The trimming filters on the output of the prism are used to shape the rising and falling edges of the spectral characteristics to improve the reproduction of neutral colors. Figure 5.3-A-2 shows an example of the spectral characteristics of a complete prism type color separation system. The color-selective characteristic of the blue-reflecting and red-reflecting dichroic coatings divert the blue and red components of the image to the respective blue and red imagers. The remaining component, green, passes straight through to the green imager. The imagers themselves have no special color characteristics. The spectral characteristics of the complete optical system are called the *taking characteristics* of the camera. The ideal taking characteristics are shown in Figure 5.3-A-3. Note that the ideal taking characteristics include negative lobes, shown dotted in the Figure. The optical system cannot generate the required negative lobes (cannot subtract light); however, subtraction of the video signals is possible in the camera processing stages. One of the functions of the linear matrix circuitry is to correct for the missing negative lobes in the optical system.

Color Separation System For Two Imager Color Cameras

The incoming light is applied to a partially silvered mirror (or prism surface) to divert half the incoming light. The straight-through portion goes to the luminance (Y) CCD imager. The reflected portion goes to the chrominance channel CCD imager (see Figure 5.3-A-4) Color separation is achieved on the chrominance imager by the use of alternate red and blue optical filters on the pixels of the CCD. The luminance resolution of the two imager system is high. Color

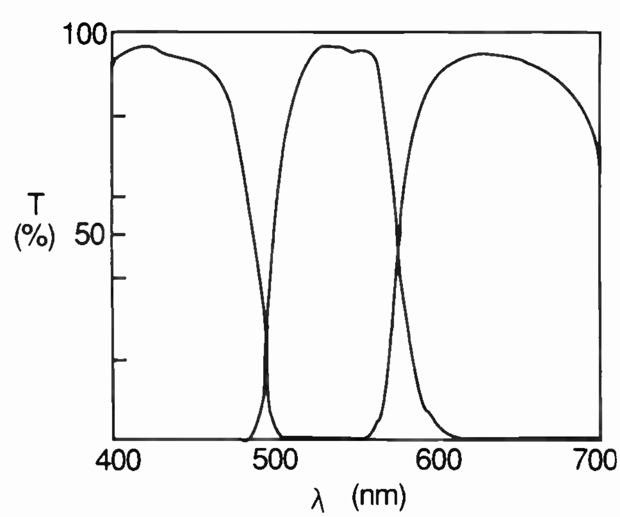


Figure 5.3-A-2. Spectral characteristics of an entire color separation system.

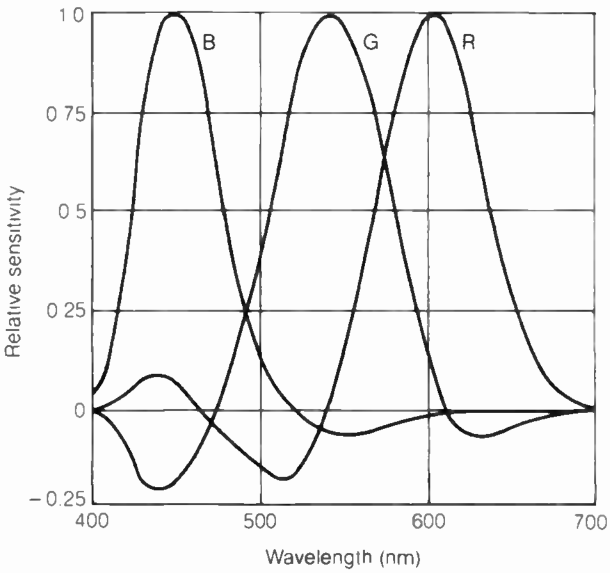


Figure 5.3-A-3. Ideal taking characteristics. *(Courtesy of Wiley Publishing.)*

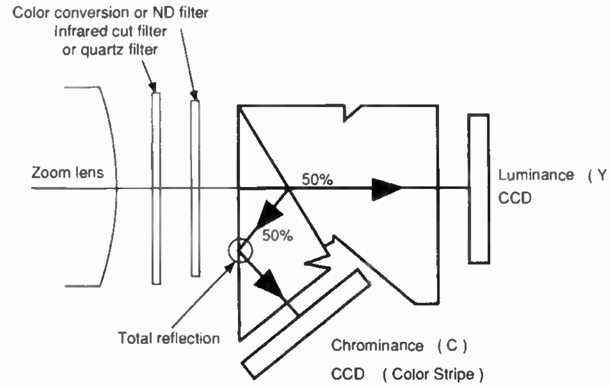


Figure 5.3-A-4 Two imager color separation system.

performance is only a little better than the one imager system.

Color Separation System For Single Imager Color Cameras

In the classic single imager camera, the red, green and blue optical filters are deposited on the surface of the CCD in a mosaic or stripe filter pattern (see Figure 5.3-A-5). The relative merits of the mosaic vs. stripe filter are given in Table 5.3-A-1. In some cases an alternate pattern of color filters is used to achieve a different combination of characteristics.

Table 5.3-A-1

Relative Performance of a mosaic vs. stripe filter.

Mosaic Filter	Stripe Filter
Higher cost	Lower cost
Higher H resolution	Lower H resolution
Higher sensitivity	Lower sensitivity
Complementary color filter	Primary color filters
Poorer colorimetry	Better colorimetry

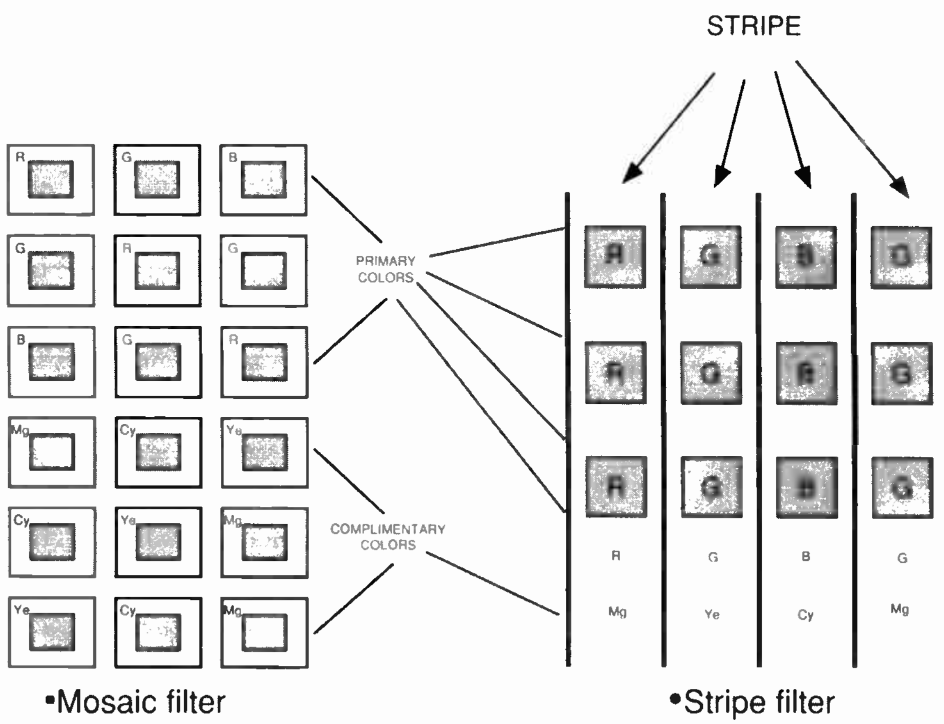


Figure 5.3-A-5. Single imager color separation system.

APPENDIX B: LENS CHARACTERISTICS

A zoom lens is typically used with a television camera. The zoom lens has the following characteristics:

- The focal length can be continuously varied over a wide range. The action of varying the focal length is called *zooming*
- The image position does not change as the lens is zoomed (focal length of the lens is changed). The ratio of the maximum focal length divided by the minimum focal length is called the *zoom ratio.*

A common type of zoom lens is illustrated in Figure 5.3-B-1. In this type of lens, a lens group (the *variator*) moves along the optical axis to vary the focal length, while another lens group (the *compensator*) is moving in a complementary manner to maintain the image in the same plane. Precisely machined cams are used to effect the complex movement of the variator and compensator elements.

Focal Length

If parallel rays of light pass through a (convex) lens, they will converge to one point on the optical axis, the *focal point* of the lens. The *focal length* of a lens is the distance from the center of the lens to the focal point. Practical camera lenses consist of several convex and concave lenses grouped together to minimize aberrations, however, this combination of lenses functions essentially the same as a single convex lens of negligible thickness (and no aberrations) also called a *thin lens* (see Figure 5.3-B-2). The focal length of a *complex lens* is then the distance from the center of this equivalent thin lens to the focal point (see Figure 5.3-B-3). The focal length is the most fundamental parameter of a lens and directly defines the angle of view of the lens.

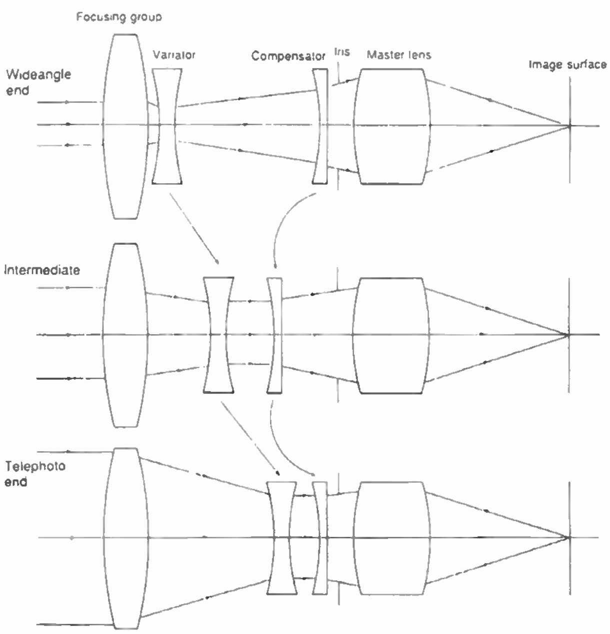


Figure 5.3-B-1. Common type of zoom lens.

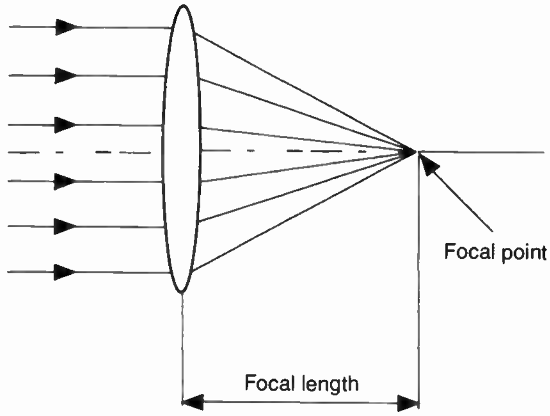


Figure 5.3-B-2. Focal length of a single or thin lens.

Angle of View

The angle of view of a lens is directly related to the image size and the focal length, as shown in Figure 5.3-B-4. Currently, the standard television systems have an aspect ratio of 4 units horizontal to 3 units vertical (4:3); therefore, the angle of view is greater in the horizontal than the vertical direction.

Field of View

The field of view is the size of an object that fills the image format (see Figure 5.3-B-5). Expressed mathematically,

$$y = 2l \tan \frac{w}{2} \quad \text{or} \quad y = y' \frac{l}{f}$$

where:

y = object dimension
l = object distance
w = angle of view
y' = image size
f = focal length

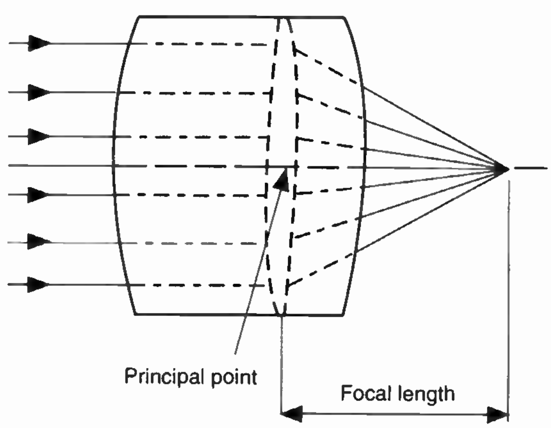
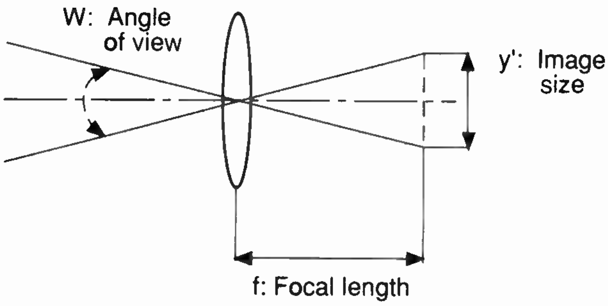


Figure 5.3-B-3. Focal length of a compound lens.



Angle of view: $w = 2 \tan^{-1} \frac{y'}{2f}$

Figure 5.3-B-4. Calculating the angle of view.

For example, if the object being viewed is 1,000 ft away and the lens has a vertical angle of view of 2°, then

Vertical field of view = 2 × 1000 × tan (2°/2)

Vertical field of view = 35 ft

Horizontal field of view
 − 35 × 4/3 (aspect ratio) = 46. 7 ft

Relative Aperture (*f*-number or *f*-stop)

This parameter characterizes the speed or light gathering power of the lens. The relative aperture is defined as f-number = f/D, where f = the focal length and D = the effective aperture diameter. A lens providing a smaller f-number is more sensitive than another lens that has a higher f-number at maximum aperture. The f-number selected is also closely related to the depth of field, where minimum depth of field corresponds to the lens at maximum aperture. For example, the typical sensitivity of a CCD type camera is 2000 lux (− 186 footcandles) at $f/8.0$ for a full output signal (see Table 5.3-B-1).

Depth of Field and Depth of Focus

Depth of field is defined as the difference between the maximum and minimum distances to a subject

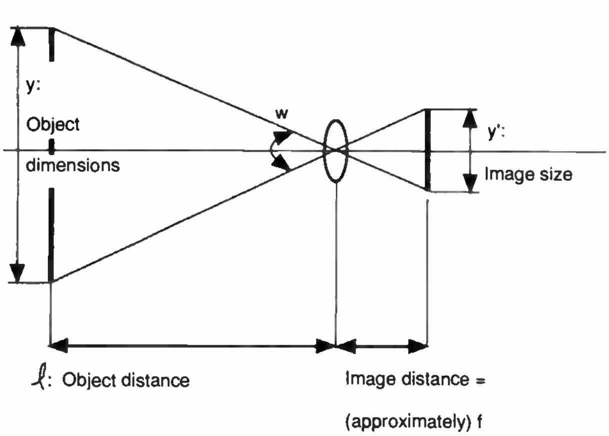


Figure 5.3-B-5. Field of view.

Table 5.3-B-1
Relationship Between f/number and Illumination Sensitivity

f/number	Illumination Sensitivity
f/8.0	2,000 lux
f/5.6	1,000 lux
f/4	500 lux
f/2.8	250 lux
f/2.0	125 lux
f/1.4	62. 5 lux

such that the image is within an allowable amount of blurring (see Figure 5.3-B-6). The human eye cannot detect blurring below the *circle of confusion* (the permissible circle of confusion becomes smaller as the resolving power of the system is increased). Correspondingly, the distance the image plane can be moved within the allowable circle of confusion is defined as the depth of focus. In CCD cameras, the CCD imagers are precisely fixed to the color separation prism and it becomes critical that the optical system be compensated such that the R, G and B images are well within the depth of focus, a difficult requirement that is now imposed on lenses intended to be used with CCDs.

The smaller the f-number of a lens, the shallower the depth of focus and therefore the smaller the depth of field. In other words, the depth of field can be controlled by adjusting the iris (aperture) of the lens in conjunction with a suitable ND filter to provide correct exposure. The depth of field characteristically becomes greater both as the focal length of the lens becomes shorter, and as the distance between the lens and the subject becomes longer (see Figure 5.3-B-6).

Ramping

Ramping or f-drop is the term applied to the drop in sensitivity (increased f-number) typically experienced in a zoom lens at the upper end of its focal length range. It is common for lens manufacturers to allow a certain amount of ramping in portable lenses to reduce the size and weight of a zoom lens. Studio lenses, however, frequently employ a large enough focusing group to avoid the ramping effect.

When the ramping effect is present in a lens, the iris must be readjusted to maintain full video level at the extreme zoom positions. In the example of Figure 5.3-B-7, the lens has a relative aperture of $f/1.6$ through a zoom range of 12 mm to 120 mm, but ramps down to $f/2.8$ at its maximum focal length of 180 mm.

Modulation Transfer Function (MTF)

MTF defines the resolution capability of a lens. MTF is typically measured by using a black-to-white sine wave pattern. The peak-to-peak amplitude of a low frequency sine wave (0.5 MHz) is used as the 100% reference; the peak-to-peak amplitude for other frequencies defines the MTF of the lens for those frequencies. Most lenses exhibit their best MTF when operated in the middle of their aperture range. For example, a

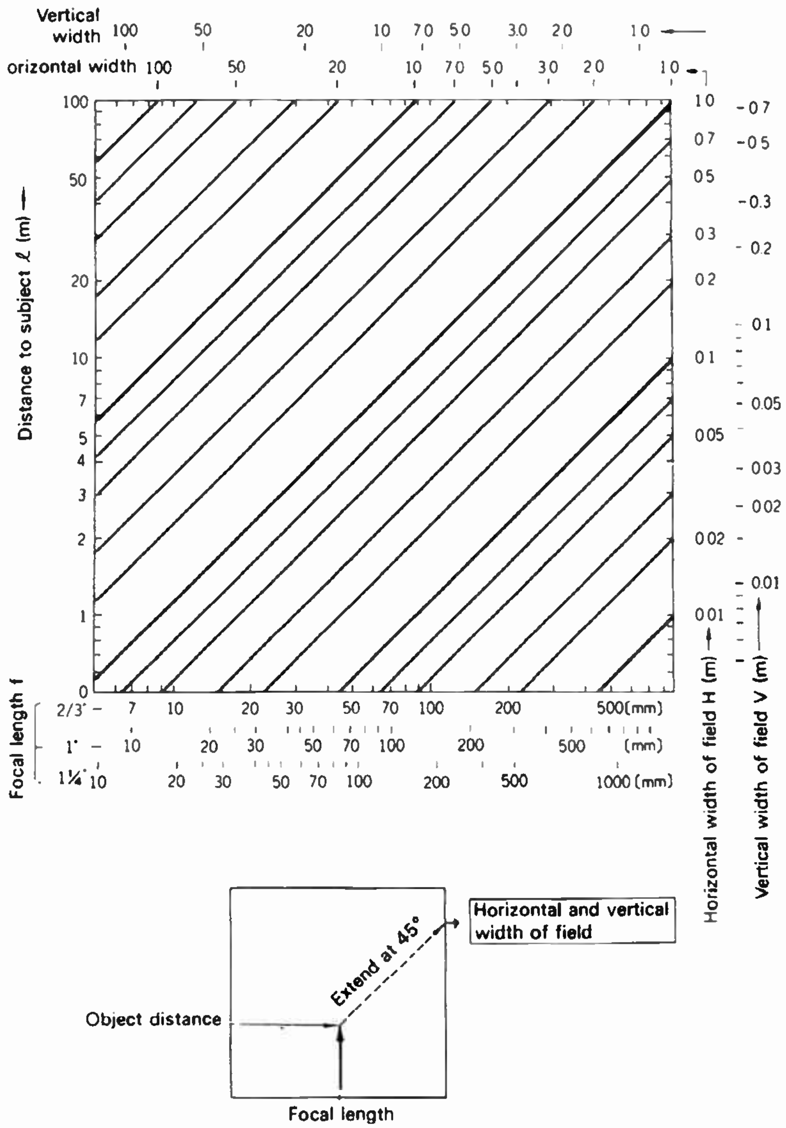


Figure 5.3-B-6. Depth of focus and depth of field.

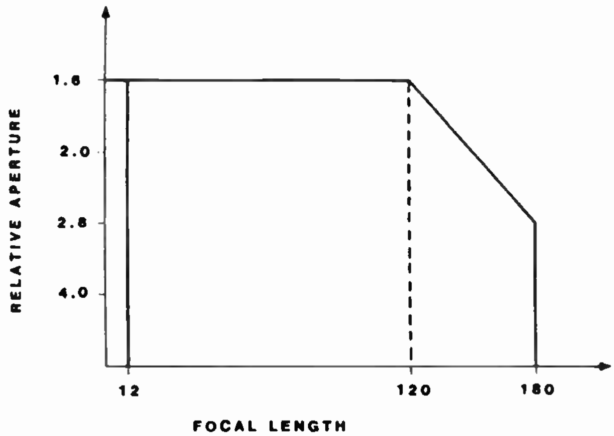


Figure 5.3-B-7. Lens ramping.

lens with an aperture range of $f/1.6$ to $f/16$ will probably gives its best MTF at about $f/4.0$. If the lens is further stopped down towards $f/16$, some small reduction in modulation depth may result because of the light being refracted as it passes through such a small aperture. However, as the lens is opened towards its maximum aperture of $f/1.6$, a more severe loss of modulation depth is likely to occur as the spherical aberrations become more significant.

In addition, the MTF of the lens tends to be best at the center of the picture and fall off toward the edges of the picture. Figure 5.3-B-8 illustrates the MTF characteristics of a typical zoom lens. Note that the MTF becomes worse as the iris is opened. This difference is more exaggerated in the corners, which illustrates the importance of considering the f-stop when evaluating the performance of a lens.

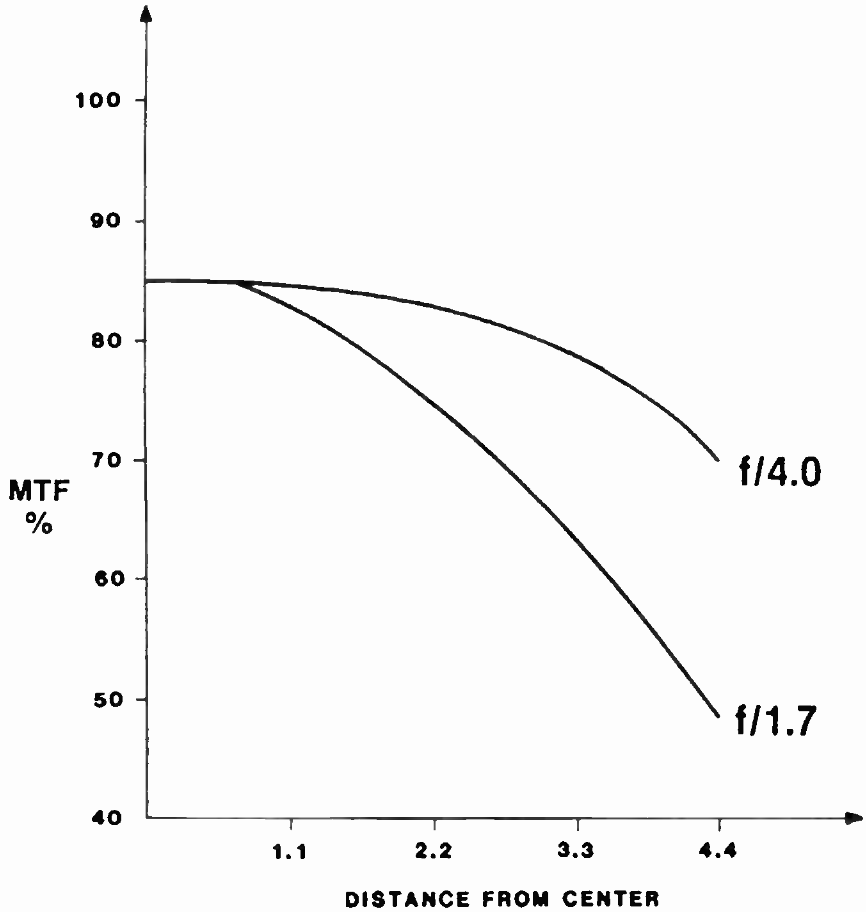


Figure 5.3-B-8. Typical MTF characteristics for 2/3 in. lens (average for all focal lengths).

Uniformity of Field (Vignetting)

Uniformity of field is a measure of the response to a white field over the full picture area. The typical uniformity of field characteristic of a zoom lens is shown in Figure 5.3-B-9, which illustrates that the response is usually greatest in the center and least in the corners. The nonuniformity in the corners becomes more pronounced at wide aperture settings and varies with focal length. High quality video cameras frequently contain a white shading circuit that increases

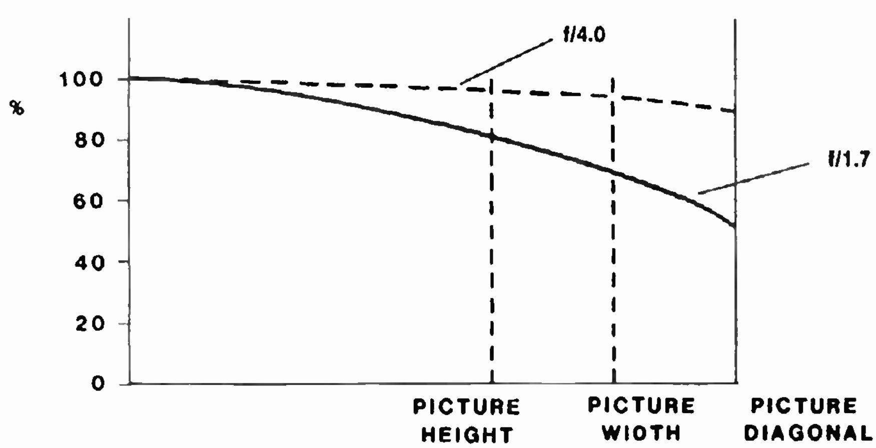


Figure 5.3-B-9. Uniformity of field.

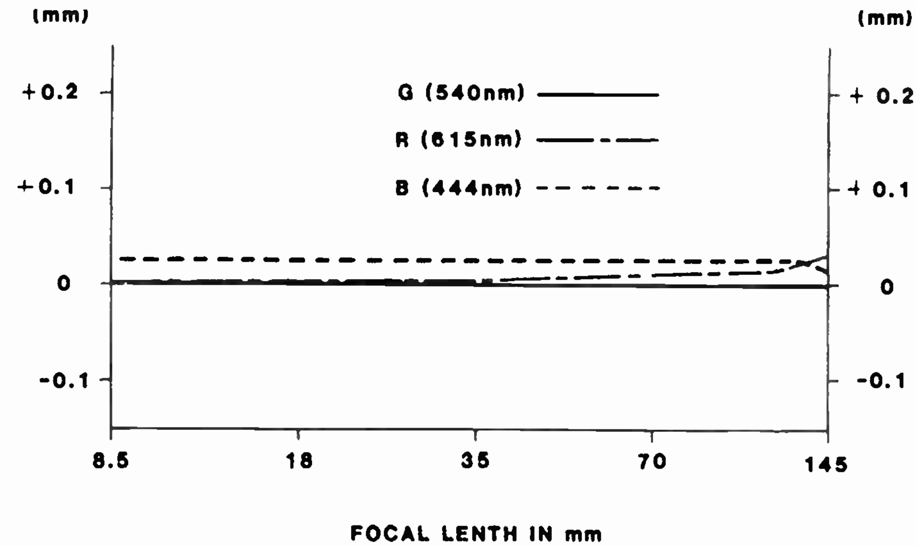


Figure 5.3-B-10. Longitudinal chromatic aberration.

gain toward the edges and provides first-order correction for this effect.

Longitudinal Chromatic Aberrations and Lateral Chromatic Aberrations

The refraction of optical glass varies with the wavelength of the light. This causes the lens to form different images for different colors (see Figure 5.3-B-10). Chromatic aberration is classified in two types:

- **Longitudinal chromatic aberration.** The position of the image plane varies according to color
- **Lateral chromatic aberration.** The size of the image on a given image plane changes according to color.

Although the lens designer makes a great effort to achieve consistency, the focal length of a zoom lens tends to vary slightly with color throughout its zoom range. This means that there will be some small differences between color channels, in both focus and image size, as the zoom is operated.

Geometric Distortion

All lenses exhibit some degree of geometric distortion, a symmetrical error (shown in Figure 5.3-B-11) affecting the red, green and blue images equally. In the case of the zoom lens, the distortion varies with the focal length such that the negative (*barrel*) distortion occurs at short focal lengths and positive (*pincushion*) distortion occurs at long focal lengths. For camera

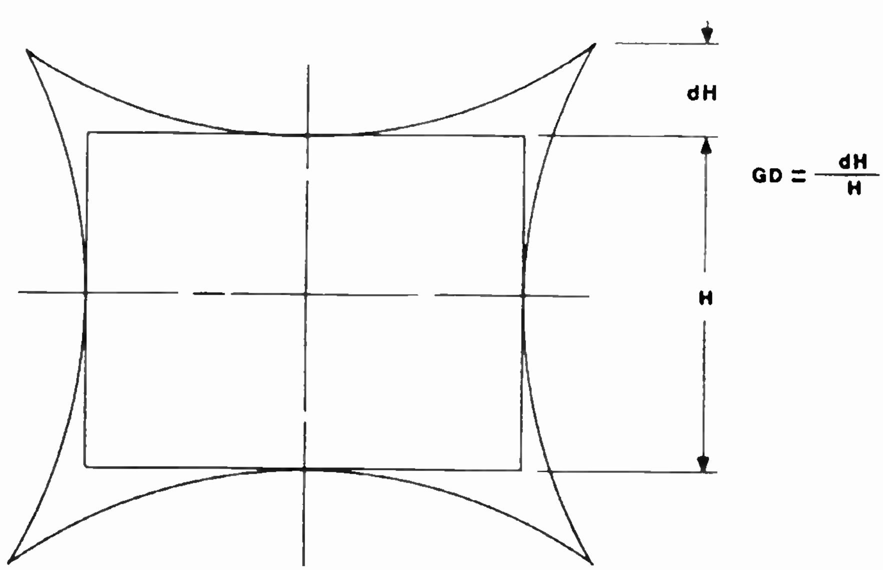


Figure 5.3-B-11. Geometric distortion.

measurements and adjustment, it is advisable to use the middle of the zoom range where this error is minimized.

Minimum Object Distance (MOD)

MOD is the closest focusing distance of a lens. Wide angle lenses generally have the shortest MOD and telephoto lenses the longest. When a lens is chosen for a specific application, it is critical to insure that the MOD is adequate.

APPENDIX C: COLOR TEMPERATURE

The illumination from an incandescent lighting fixture is frequently described as a warm source of light by cinematographers, and indeed, when analyzed with a spectra device that measures the relative intensity of the color components, red wavelengths described as a warm color, predominate. Daylight illumination on the other hand is frequently described as being cool and the same spectral analysis shows blue wavelengths, usually described as a cool color, to predominate. The concept of color temperature makes it possible to describe the character of the illumination in a much more precise manner than these vague descriptions. The concept of color temperature also makes it possible to easily determine the proper color correction filters to use with a specific source of illumination.

Color temperature is defined as the visible spectrum emitted by a black body that has been heated to specific temperature in degrees Kelvin. The color of an illuminant is determined only by its temperature and is almost the same for all substances at a given temperature. For example a quartz halogen studio light is said to have a color temperature of 3,200°, because the visible spectrum of the studio light closely resembles the spectrum emitted by a black body heated to 3,200° K. Daylight is said to have a nominal color temperature of 5,600°,

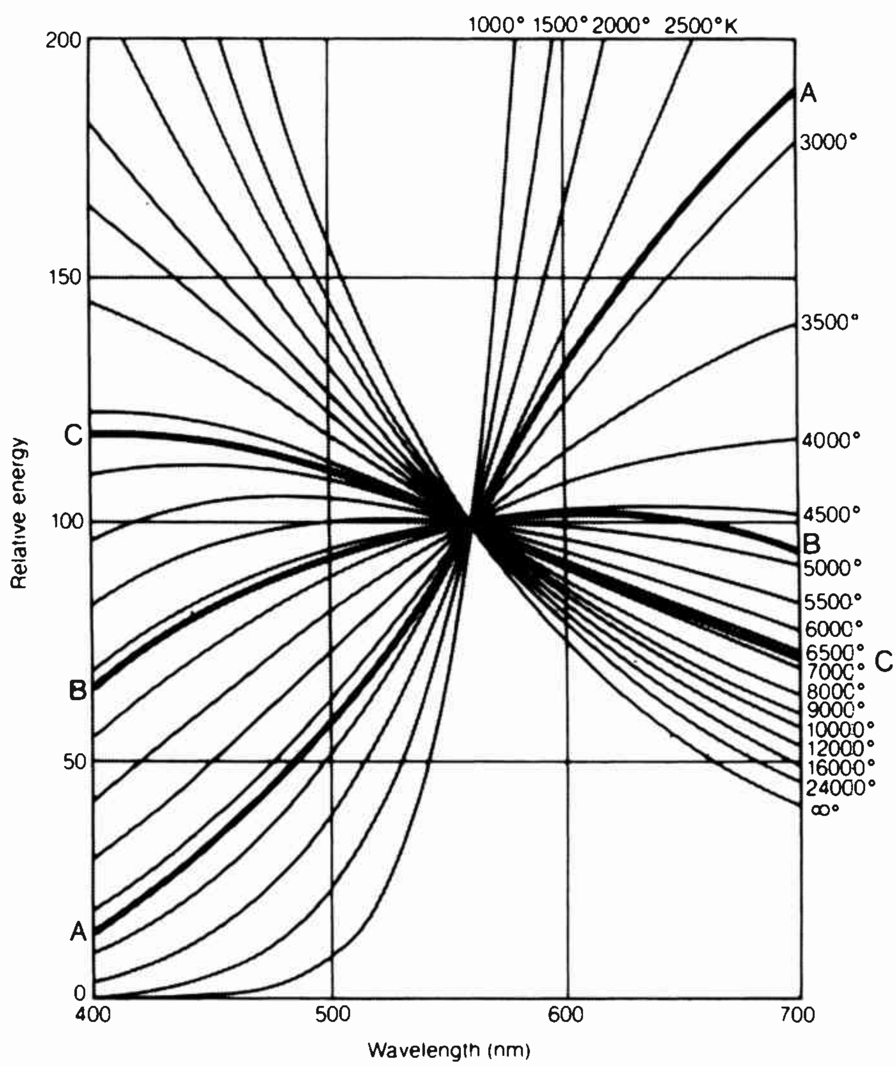


Figure 5.3-C-1. Spectral distribution at each color temperature.

because its spectrum provides a close match to the spectrum of a black body heated to a temperature of 5,600° K.

The light emitted by a black body that has been heated, is a smooth, continuous spectrum that covers the complete visible spectrum. Lighting sources such as metal vapor lights and fluorescent lights typically have a discontinuous spectrum of strong spectral lines. Nevertheless, a color temperature designation is frequently assigned to such sources, to give an overall indication of the character of the light as perceived by the eye. These lighting sources, frequently used in stadiums and arenas because of their efficiency, must be used with great caution to achieve acceptable color reproduction with a color television system.

The relative energy vs. wavelength at various color temperatures is given in Figure 5.3-C-1. For example, consider an illuminant of approximately 2,900° K (curve A). The magnitude of the long wavelengths which correspond to red (600–700 nm wavelength) are large, while the magnitude of the short wavelengths corresponding to blue (400–500 nm wavelength) are very small. This relationship reverses for high color temperatures.

APPENDIX D:11-STEP GRAY SCALE

An 11-step gray scale chart is highly recommended for use in adjusting a camera. The 11-step chart incorporates a true gamma of 0.45, as well as a 40:1 contrast ratio. It is highly recommended not to use the traditional EIA 9-step logarithmic gray scale chart for camera verification and adjustment for the following reasons:

- All the instruction manuals provided with current cameras provide waveforms and voltages appropriate for an 11-step chart. It is extremely difficult to relate these waveforms and voltages to a 9-step chart.
- The EIA chart uses logarithmic steps that do not mathematically fit the gamma curve of the camera
- The 9-step chart covers only a 20:1 contrast ratio, inadequate to test a modern video camera that is capable of at least a 40:1 contrast ratio.

REFERENCES

1. Benson, K. B., and Jerry C. Whitaker, (eds.), *Television Engineering Handbook,* revised ed., McGraw-Hill, New York, 1992.
2. Crutchfield, E. B., (ed.), *NAB Engineering Handbook,* 8th ed., National Association of Broadcasters, Washington, D.C., 1992.
3. *TV Optics—The Canon Guidebook of Optics for Television Systems,* Broadcast Equipment Center, Canon, Inc., Kawasaki, Japan, 1986.

The author wishes to express his sincere thanks to Mr. Larry Thorpe and all the members of the Sony team that have contributed valuable material and assistance in preparing this chapter.

5.4 VIDEO SIGNAL SWITCHING, TIMING AND DISTRIBUTION

JAY KUCA
GRAHAM-PATTEN SYSTEMS, GRASS VALLEY, CA

INTRODUCTION

Evolutionary changes in video equipment technology have had a dramatic impact on the design of video systems. Not that long ago, composite analog systems were the norm. Today, however, the typical facility is a hybrid of composite and component, analog and digital systems.

Engineers and technicians working in this mixed format environment must be aware of the system design considerations that are unique to each format. They must also have a clear understanding of the implications of converting between formats. There are four basic video signal formats which are currently in use in professional video installations: composite analog (NTSC), component analog (RGB or Y, P_r,P_b), composite digital (4fsc), and component digital (Rec. 601).

Most of the infrastructure which is in place in television stations today is composite analog. This is because NTSC has been in existence longer than the other formats and a substantial amount of composite analog equipment is still in service. It is also due to the fact that the current terrestrial delivery system is NTSC based.

Newly acquired equipment is more likely to be digital than analog. As the balance of analog versus digital equipment in a facility shifts more toward digital, conversion of the facility's routing and distribution systems will become more attractive. This will be accelerated by the conversion of the terrestrial delivery system to digital.

NTSC and digital television (DTV) are likely to coexist for many years to come. Throughout the transition, it is important that technical staff maintain proficiency in both analog and digital technology.

NTSC SYSTEMS

In 1953, the National Television Systems Committee (NTSC) recommended a set of color television standards to the Federal Communications Commission (FCC). The FCC adopted the standards and color television transmissions commenced shortly thereafter.

These color standards were based on the monochrome standards developed by the NTSC in the early 1940's. In fact, a key requirement of the color standard was that the transmitted signal be compatible with existing monochrome receivers. This was accomplished by retaining all of the characteristics of the monochrome signal and adding color information through the use of a quadrature modulated subcarrier added to the monochrome signal.

The subcarrier frequency was carefully chosen to fit within the channel bandwidth to minimize its visibility on the screens of monochrome receivers and interference with the 4.5 MHz sound carrier used in transmission. The actual frequency which was selected is 3.579545 MHz ± 10Hz. It is commonly referred to as 3.58 MHz.

NTSC Video Signal Characteristics

Figure 5.4-1 is an oscillogram of a typical video signal. The peak to peak amplitude of the signal is 1 V peak to peak or 140 IRE units. This is the standard level for signal distribution in all professional television facilities. The signal is comprised of +100 IRE units (714 mV) of picture and −40 IRE units (286 mV) of sync.

Even though video is an ac signal, proper operation of most equipment depends upon a fixed dc component. By convention, blanking (0 IRE) is assumed to be at ground (0 VDC). This reference is lost, of course, in portions of the system which are ac coupled, but when the signal is clamped or dc restored, standard practice dictates re-establishing blanking at ground.

The bandwidth of a studio quality NTSC video system is determined primarily by the state of the art and economic factors. And while it can be argued that the transmission system cuts off any information above 4.2 MHz, it is good engineering practice to strive for response which is flat to 8 to 10 MHz in the studio.

Timing integrity is ensured through the use of a *master sync pulse generator*. In nearly all installations, timing for all signals in the plant is derived from this single reference. In most cases, the master sync generator is driven from a high stability crystal oscillator. Occasionally, a *precision rubidium* or *cesium reference oscillator* is used. (Figure 5.4-2 illustrates the relationships between the horizontal, vertical and color timing signals.)

Interconnection Conventions

All analog composite studio equipment is designed to be interconnected with coaxial cable with a nominal characteristic impedance of 75 U. In the simplest case, a point-to-point connection between two pieces of equipment, a continuous length of cable is driven from a 75 Ω source and terminated in a 75 Ω load (see Figure 5.4-3).

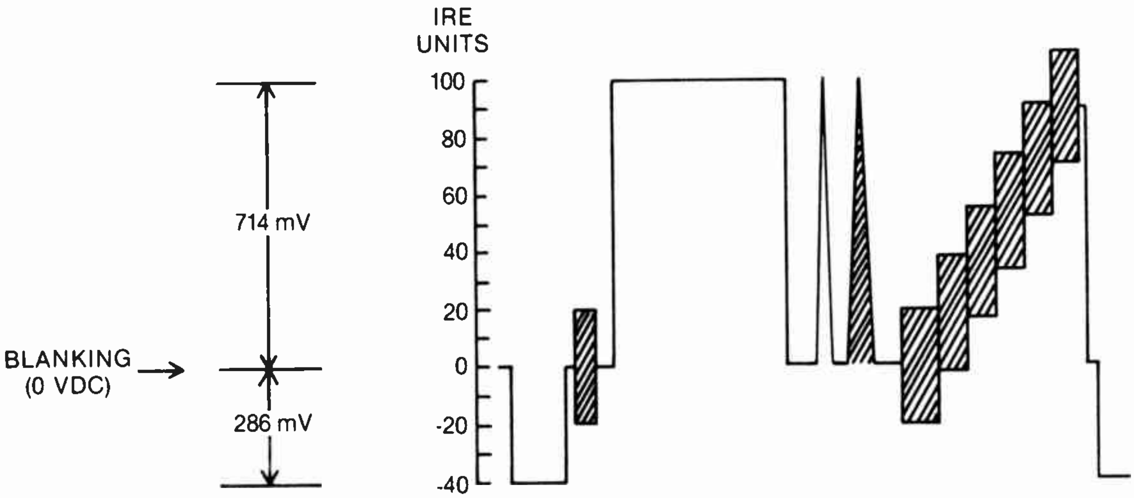


Figure 5.4-1. Oscillogram of a typical video signal.

Two possible approaches exist when it is necessary to distribute the signal from a single source to more than one destination. The first is entirely passive in nature: one end of the cable is driven from the 75 Ω source and, instead of terminating the far end in a 75 Ω load, a *loop through* connection is made to the first piece of equipment. The loop through is carried on to the next piece of equipment, and so on. A 75 Ω termination is placed on the loop through connection on the last piece of equipment. This approach, shown in Figure 5.4-4, will work well provided the loop through inputs are properly designed and the cable lengths are kept short. The number of loop throughs should also be kept as small as possible to prevent frequency response errors and, in severe cases, signal reflections.

Terminations should be made using *non-inductive precision resistors.* An ideal termination should present a purely resistive load of exactly 75 Ω. Any reactive components will cause frequency response errors. Errors in the value of the resistive component will cause errors in signal levels. Terminations made with 0.1% tolerance metal film resistors are recommended.

Whenever possible, video signals should be distributed using an active device such as a *distribution amplifier or routing switcher.* This effectively results in the equipment being interconnected in the point-to-point manner previously described.

Nearly all modern equipment uses BNC connectors for video input and output connections. Some older equipment still in service may use UHF (PL-259/SO-239) connectors. Occasionally, space limitations dictate the use of a subminiature connector such as a BSM series. However, the BNC is by far the most common connector in use.

A wide variety of cable types are available which are suitable for high quality video interconnect. One popular type is Belden 8281, a double shielded 75 Ω precision coaxial cable designed specifically for video use. 8281 provides a good balance between loss and physical size. Its double shielding helps reduce the likelihood of stray signal pickup.

Where space limitations, increased flexibility, or other factors require the use of a smaller diameter cable than 8281, an RG/59 type cable is the usual choice. In choosing such a cable, make certain it has a solid copper center conductor and braided copper shield. There are many RG/59 type cables on the market with copper clad inner conductors and foil shields. These are designed for RF applications such as CATV and are not suitable for use in baseband video systems.

Distribution Amplifiers

The distribution amplifier or (DA) is the most elementary building block in a video system. Its primary function is to allow a single 75 Ω video source to feed multiple 75 Ω loads with no apparent loss. Quite often, other functions such as cable equalization are included in the DA's circuitry.

Even though DA functions and designs vary, it is quite easy to construct a generalized model. As shown in Figure 5.4-5, the typical DA consists of an input amplifier followed by a conditioning stage (gain, equalization, delay, etc.), followed by an output amplifier.

The input amplifier is generally designed with a rather high input impedance (several tens of kilohms). This allows the driving signal to be looped through the input with minimal loading effects. Quite often, a complex network is used at the input connection to cancel out any reactive components which might affect the loop through. The input stage may be either single ended or differential, the latter being more common on high quality DA's.

A differential input stage can be very useful in combating one of the most serious problems encountered in unbalanced signal transmission—common mode hum and noise. In larger facilities, common mode hum due to ground loops may run as high as several volts. Even in small plants with excellent ground systems it is

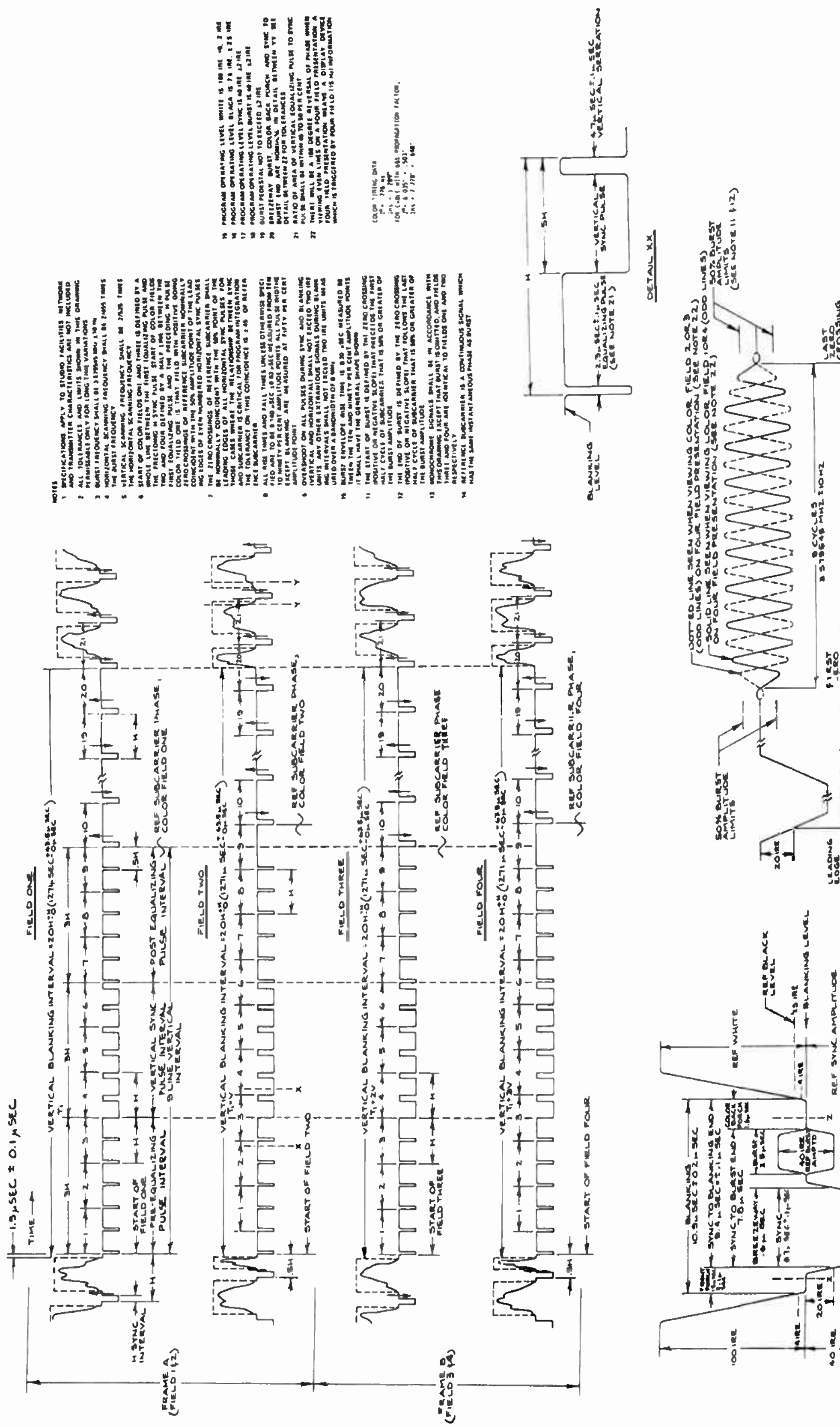


Figure 5.4-2. NTSC timing parameters. An updated version of the EIA Tentative Standard RS-170A. *(Courtesy of the Grass Valley Group)*

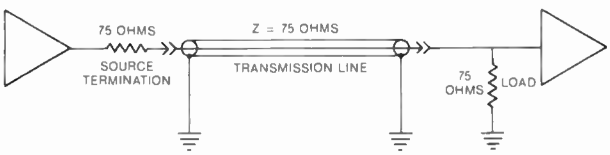


Figure 5.4-3. Point-to-point video connection.

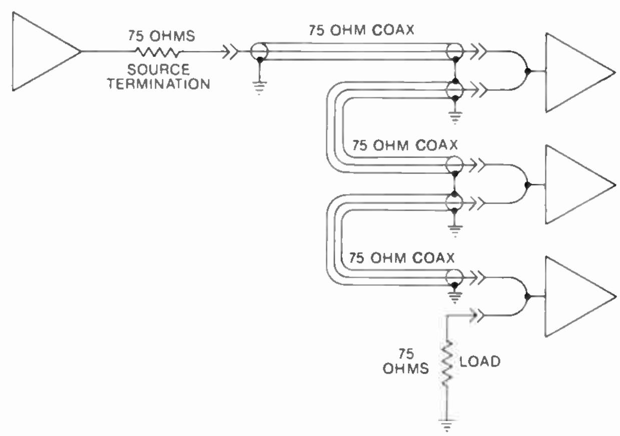


Figure 5.4-4. Loop through video connection.

possible to have several tens of millivolts of difference in ground potential between different pieces of equipment. A differential amplifier will cancel any common mode signals while amplifying the (desired) differential mode signal. Common mode rejection ratios (CMRR) of 50 to 60 dB are typical of most modern differential input DA's.

The conditioning stage may be a *wideband gain stage*, *active equalizer stage*, or *delay amplifier*, depending upon what additional function the DA performs. If distribution is the only function, the conditioning stage may be omitted.

The *output stage* is an amplifier with very low output impedance. It is followed by source termination resistors which ensure that the cable recognizes 75 Ω at the sending end as well as the load end. High output to output isolation is ensured by the low impedance of the output stage.

Cable equalization is often required in video systems. This is due to losses increasing with frequency in coaxial cable. Belden 8281 has approximately 0.8dB of loss per 100 ft at 10 MHz. Since runs of 100 ft or more are not uncommon in most facilities, achieving flat system response makes equalization of these losses necessary. The approach taken by most manufacturers is to construct an equalizer network with a response that closely approximates the inverse of the cable loss curve. When this equalization is applied to the cable/DA system, the result is flat overall response.

Cable equalizers must be designed for a specific type and length of cable. When a variable equalization control is provided, it usually serves to adjust the amount of equalization applied. Because cable loss is a complex function, the equalization curve is only correct for one cable length. Slight response errors will be noted at other length settings.

Cable losses should always be equalized at the load end of the cable. The reason for this is simple: if a flat signal is applied to the input of the amplifier, the high frequency boost of the equalizer stage will be added to the signal and the output stage will have to track these larger than normal amplitude high frequencies. In cases where the amount of equalization is more than a few dB, the output stage will be unable to deliver the necessary current into its low impedance load. If, however, the DA is located at the load end of the cable, the cable losses have attenuated the high frequencies at the input of the amplifier. The high frequency boost of the equalizer stage restores the flat response characteristic to the signal, and the output amplifier does not have to deal with larger than normal swings at high frequencies.

Another function which is often required in video systems is *clamping*. Clamping is used primarily to reduce low frequency hum and tilt but may also be used to restore the dc component of a video signal which has passed through one or more ac coupled devices. Clamps operate on a line-by-line basis and, as such, are fast acting. They can be triggered by noise impulses and should therefore not be used on noisy signals.

Direct current restoration may be used in lieu of clamping to reduce the dc signal bounce due to multiple

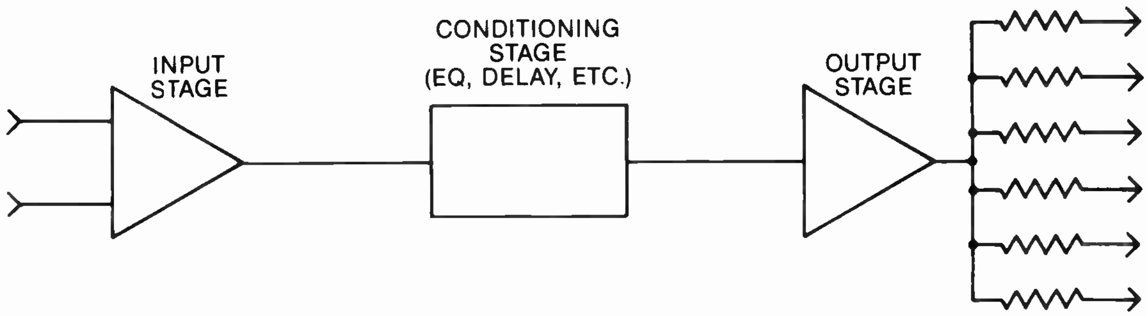


Figure 5.4-5. Simplified block diagram of a video DA.

ac couplings. Because of the slow time constants employed in dc restorers, they are not affected by noise; but by the same token, they are not useful in eliminating hum and tilt.

Delay DA's are often used in place of coaxial cable or other passive delay elements. Delays of up to a microsecond or more are possible on modules no larger than a conventional DA. Many delay DA's include circuits to compensate for delay line response errors which produce better results than a passive delay line and DA in combination. Some delay DA's also include other functions such as cable equalization.

Distribution Amplifier Selection

As a signal traverses a typical path through a typical system, it may pass through ten or more DA's. Therefore, the performance of a single amplifier must be, for many parameters, an order of magnitude better than the system specification. Table 5.4-1 lists the performance specifications of a representative unit.

Selecting the right DA for a particular application involves determining the desired level of performance and the features and functions required. Since most manufacturers offer a range of models which all fit into the same mounting tray, it is wise to look at the sum total of all requirements and then choose the manufacturer who can best accommodate those requirements.

DA's fall into the *install it and forget it* category—low maintenance and long service are important factors. When selecting distribution amplifiers, consider the quality of the components which the manufacturer has used. Will they hold up for ten or fifteen years? Is the design a conservative one, or are devices being run at their limits? Thermal stress will not only lead to equipment failure, it will cause drift as well.

Pulse and Subcarrier Distribution Amplifiers

Black is by far the most commonly used signal for equipment synchronization. However, separate pulses and subcarrier are still used as timing references in some systems. *Pulses* (sync, blanking, H drive, V drive) are generally distributed as 2 to 4 V negative going signals. Subcarrier is generally distributed as a 1 to 2 V peak-to-peak signal.

Table 5.4-1
Typical video DA performance specifications

Frequency Response	+/− 0.025 dB to 5 MHz, −0.2 dB at 10 MHz
Differential Phase	<0.1°
Differential; Gain	<0.1%
T-Pulse to Bar	<1.0%
Tilt	<0.5%
Chrominance/Luminance Delay	< 10ns
Hum and Noise	>70 dB below 1 V p-p
Common Mode Rejection	>60 dB
Input Return Loss	>40 dB to 5 MHz
Output Return Loss	>40 dB to 5 MHz
Output Isolation	>40 dB to 5 MHz

Most video DA's will not accommodate signals larger than 2 volts peak-to-peak. So, while most video DA's would be suitable for subcarrier distribution, the typical video DA will not handle full amplitude pulses. DA's designed specifically for pulse distribution are optimized for large negative signal swings. They will not function well, if at all, as video amplifiers.

There are two general categories of pulse DA's—regenerative and linear. Regenerative DA's include circuitry for pulse regeneration, typically edge detectors driving one-shots. Timing of the one-shots is usually adjustable, allowing the relative delay of the amplifier to be varied. Delays of up to several microseconds are common. Due to the inherent delay of the pulse regeneration circuitry, these amplifiers always exhibit some value of minimum delay.

Regenerative subcarrier DA's are also available. These amplifiers generally feature adjustable circuits to shift the phase of the subcarrier. The adjustment range is usually slightly more than 360.

When pulses and subcarrier are distributed through regenerative DA's, there is a good chance that the subcarrier to horizontal (SC/H) phase will not be maintained throughout the system. This is because it is possible to adjust sync and subcarrier timing independently. For this reason, linear (non-regenerative) pulse and subcarrier DA's should be used in SC/H phased installations. Specially designed linear pulse DA's are available and most high quality video DA's can be used as linear subcarrier DA's.

Video Signal Switching

Analog video signals can be switched using a variety of methods. Simple monitor selectors may be nothing more than passive mechanical switches. In some cases, electromechanical devices such as relays are employed. Most modern switching systems employ solid state switching devices.

Figure 5.4-6 is a simplified block diagram of a video switching system. Input amplifiers buffer the input signals and provide drive to the switching elements (crosspoints). The crosspoints are bussed in horizontal

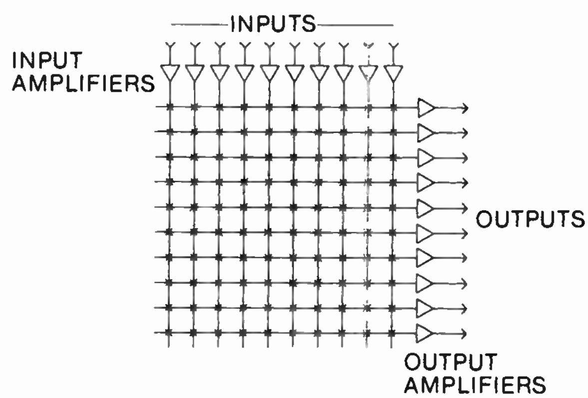


Figure 5.4-6. Simplified block diagram of a video switching system.

rows (output buses) to an output amplifier. If the switcher has more than one output bus, the buffered input signals are bussed in vertical columns (input buses) which run across the output rows.

Not shown in the block diagram is the crosspoint control logic. In single bus switchers with ten inputs or less, control is generally accomplished using a single wire per crosspoint. In single bus switchers with more than ten inputs, an encoding scheme such as BCD is generally used to reduce the number of wires required. Larger multiple bus switchers generally use some sort of serial control scheme.

Single Bus Switchers

Single bus switchers run the gamut from simple to elegant. For noncritical applications such as picture monitor input selection, a mechanical switcher may be adequate. Most mechanical designs are entirely passive; that is, they have no input or output amplifiers. As such, they have terminating inputs and require dedicated (non-loop through) input feeds.

A slightly more sophisticated design adds an output amplifier only. Switching is still done using mechanical switches, but the bus is made high impedance to allow looping inputs. This approach makes it possible to stack units to provide an economical multiple input/ multiple output configuration. In terms of performance, this type of switcher is no better than the passive type. If the input loop throughs are not compensated, its performance may actually be worse.

Any single bus switcher that is in a program path should be designed more along the lines of Figure 5.4-6. In other words, it should have input buffer amplifiers, solid state crosspoints and an output amplifier. If it is being used for live switching, the buffer amplifiers should clamp or dc restore the input signals to provide bounce free switching. To minimize disturbances in the output signal when switching between synchronous inputs, the design should provide vertical interval switching. If the unit is used in an actual program loop, the design should permit servicing with minimum disruption; that is, all active circuitry should be mounted on plug-in printed circuit boards.

Single bus switchers can be combined to make multiple bus switchers, but in most cases the result will cost more and lack the flexibility and performance of a true multiple bus design.

Multiple Bus Switching Systems

There are two basic architectures for multiple bus switching systems. One is *output oriented switcher* in which the circuitry for each bus is physically separate from the circuitry associated with other buses in the system. The other is a *matrix switcher* in which the circuitry for multiple output buses is packaged together.

Output oriented switchers occupy more rack space and are more expensive than comparable matrix switchers because it requires more duplication of circuitry and more interconnect. The chief advantage of an output oriented switcher over a matrix switcher is that circuit cards can be removed from an output oriented switcher without disturbing multiple signal paths. This is an important consideration in large facilities such as network switching centers.

Matrix architected switchers are physically more compact and less expensive than comparable output oriented switchers. They offer several other advantages as well: reduced power consumption, slightly higher reliability due to fewer mechanical interconnections and shorter signal path. Matrix switchers are generally designed around a relatively small square or rectangular building block, making input and output expansion in bite sized increments relatively simple.

Input signal distribution in both types of switchers can be handled using either looping techniques or distribution amplifiers. The cost objectives of most matrix designs suggests the use of looping while the large physical size of the output oriented switcher makes the distribution amplifier approach almost a necessity.

As is the case with single bus switchers, if live switches are made on a multiple bus switcher, the input signals must be clamped or dc restored to ensure transient free switching. This signal conditioning may be done externally, but better overall performance will result if it is done in the switching system. The reason for this is that a common precision dc reference can be used for all inputs if the clamping or dc restoration is done inside the system.

Switch timing is also important if glitches are to be avoided. Most multiple bus switching systems have a sync input which is used as a reference for a master trigger pulse generator. The trigger is set to occur during the vertical interval, thus ensuring glitch free switching between synchronous sources.

Timing scatter is an important parameter for any switching system that will be used in a studio environment. It is the measure of the delay difference between paths in the switcher. In a system which uses looping techniques to distribute the input signals, there will be a fixed amount of delay between output bus groups. But on a given bus, there should be very little path length difference from input-to-input: less than plus or minus 1 of subcarrier is a typical figure. In a system that uses DA's to distribute the input signals, it should be possible to achieve the same timing accuracy from any input to any output. Representative performance specifications that should be expected of a modern routing switcher are listed in Table 5.4-2.

With respect to the use of routing switchers in video facilities, two basic schools of thought prevail: a single large switcher as a master grid, or a network of smaller switchers distributed throughout the facility. With the master grid approach, one large switching system is used with all switched sources in the plant appearing as inputs. This provides maximum flexibility, but at considerable expense. If all sources are not needed at all destinations, many of the crosspoints in the system will not be utilized.

In the distributed approach, the plant is subdivided into functional areas (studios, master control, transmis-

Table 5.4-2
Typical video switching system performance specifications

Frequency Response	±0.1 dB to 5 MHz, −0.5 dB at 8 MHz
Differential Phase	<0.1°
Differential Gain	<0.1%
2 T-Pulse to Bar	<0.25%
Tilt	<0.5%
Crosstalk	>60 dB to 5 MHz
Hum and Noise	>70 dB below 1 V p-p
Input Return Loss	>40 dB to 5 MHz
Output Return Loss	>40 dB to 5 MHz
Switching Transients	<30 mV
Timing Scatter	±1° maximum, input to input on any bus

sion, central machine room, news editing, etc.) and a small switching system placed in each area. Each system has only those area sources it requires plus a few master sources common to the other areas (color black, bars, etc.) Tie lines between matrices provide connectivity between functional areas.

The distributed approach can be considerably more cost effective than the master grid approach. However, careful consideration must be given to overall system design and tie line management. A sufficient number of tie lines must be provided to accommodate worst case connectivity needs, otherwise, system users might experience busy signals, an unacceptable condition in most broadcast facilities. And in all but the smallest, simplest systems, tie line management should be relegated to a computer. In the final analysis, the complexity and potential limitations of a distributed system with tie lines must be weighed against the higher cost of a large master grid system.

Switching System Control

Microprocessor based control systems have enhanced the operational capabilities of multiple bus routing switchers. Microprocessor control makes it possible to select sources using their familiar names (BLACK, BARS, or VTR 5) instead of input numbers. Many systems allow the naming of destinations (STL, UPLINK, or VTR 12) and levels (VIDEO, AUDIO 1, AUDIO 2, or KEY) as well. Some systems provide advanced control features like salvo switching, privileged access, and virtual matrix mapping.

These enhanced capabilities add considerable complexity to the control system. To minimize downtime due to control system failures, it is advisable that any routing switcher in an online environment be equipped with dual control computers with a fail-safe changeover system.

Video Signal Timing

Timing requirements stem from the fact that a video signal is a serial stream of analog information which is used to electronically reconstruct a picture. Horizontal timing information is used to trace out each line of the picture on the face of a display tube in exact synchronism with the camera tube beam which scanned the original scene. Vertical timing information is used to signal the end of each still picture in the frame-by-frame reconstruction process. Color timing information in the form of a brief burst of subcarrier each line is used to ensure that the color at each point on the screen is accurately reproduced.

Timing in the television studio is the responsibility of a ubiquitous black box called a *sync generator.* In the early days, a monochrome sync generator filled a 6 ft equipment rack. Today, color units which are several orders of magnitude more accurate and stable can be packaged on a single printed circuit card.

The size and complexity of the early generators prompted the master generator approach to plant timing. Since it would have been impractical to include a sync generator with each source, a set of signals was defined that would provide the information necessary to generate and process a television picture. The responsibility for generating these signals fell to the master sync generator. These signals were then distributed to each piece of equipment requiring them. They included sync and blanking (used mainly in video circuits) and horizontal and vertical drives (used to drive the sweep circuits in cameras).

The introduction of color meant that a 3.58 MHz subcarrier had to be generated by the same master generator and distributed throughout the plant. In addition, another pulse called a burst flag was needed to gate the color burst. Six separate signals, just to keep everything locked to the same timing reference.

The advent of solid state devices led to a dramatic reduction in the size of the master sync generator. At the same time, integrated circuit logic made it possible to easily derive horizontal and vertical drives and burst flag from sync. Before long, timing input requirements dwindled to sync, blanking and subcarrier on most equipment.

The capabilities of large scale integrated circuit (LSI) technology has made it possible to reduce most of the circuitry required for pulse generation to a single chip. For this reason, most modern equipment has its own sync generator built in. In addition to making stand-alone operation possible, this reduces the input timing reference requirements to a single wire reference such as color black.

Figure 5.4-7 is a simplified functional block diagram of a sync generator. A precision oscillator is used to

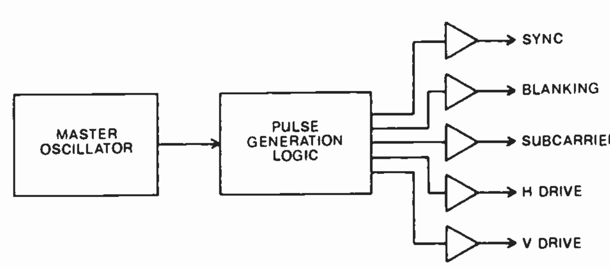


Figure 5.4-7. Simplified block diagram of a basic sync generator.

drive a set of complex logic which generates all of the required timing signals. Buffer amplifiers convert the logic levels to the required 4 V negative going pulses and 2 V peak-to-peak subcarrier.

The FCC requires that subcarrier frequency be maintained within 5 Hz of 3.579545 MHz. Therefore, an ovenized or temperature compensated crystal oscillator is generally employed.

Most modern sync generators utilize a single LSI integrated circuit as the heart of the pulse generation logic. Numerous off the shelf and custom parts exist which have 90% or more of the logic required. External parts are often needed to provide additional features such as variable pulse widths.

All IC logic families produce pulses with very fast rise times. So, in addition to providing the proper impedance and voltage levels, the output buffer stages often include shaping or filtering circuits to control the rise times of the output pulses. Controlled rise times help to ensure the reliable and predictable operation of pulse detection and regeneration circuits in the equipment being driven by the sync generator.

Genlock

A considerable amount of additional circuitry is required to enable a sync generator to lock to a composite video signal such as color black. Figure 5.4-8 shows the basic functional blocks: an input amplifier, subcarrier regenerator, sync separator and video presence detector.

The input locking signal may contain common mode hum or noise which could upset the operation of the circuits which follow. For this reason, a differential input amplifier is generally used, followed by a clamp. Since the locking signal may not always be present, the genlock circuitry usually includes a video presence detector. The output of this detector is a logic level signal which tells the sync generator whether to select the *free run* or *locked* mode of operation.

The buffered and clamped input signal is low pass filtered and then fed to a sync separator which outputs sync and burst gate. The burst gate, along with high pass filtered video, is used to regenerate the 3.58 MHz subcarrier present on the input locking signal. The separated sync and subcarrier drive phase locked loops in the generator. Adjustments of these locking circuits permit the generator phase to be varied with respect to the phase of the input signal.

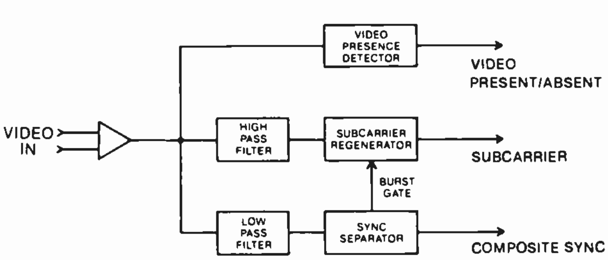


Figure 5.4-8. Simplified block diagram of basic sync generator genlock circuits.

The generator may be genlocked to the input signal in one of two basic ways: by independent locking of sync and subcarrier or dependent locking of these two signal components. With the independent method, the generator's sync and subcarrier locking circuits are continuously referenced to the input signal's sync and subcarrier. With the dependent method, lock is initiated in the same manner as the independent method. But once full genlock is achieved, the input signal's subcarrier is used as a reference for both loops.

An independent locking generator will track the SC/H phase of the input signal. It will also lock to a non-standard color signal (one in which the subcarrier and sync are not phase locked at all). A dependent locking generator will maintain correctly SC/H phased outputs regardless of the SC/H phase of the input signal. It will also lock more reliably to signals with occasional noise hits or with sync-only timebase error.

The once common practice of genlocking the station's master generator to a non-synchronous source such as network has given way to the use of frame synchronizers. However, genlocking is still used in master/slave sync systems where individual sources or groups of sources are driven off slave generators which are locked to the plant's master generator. And, as was previously mentioned, many of today's cameras and VTR's contain their own sync generators which are locked to an external color black reference.

As a result, it is not uncommon to find situations where three or more sync generators are cascaded. This makes it very important that the first generator in the line (usually the plant's master generator) has very good timebase stability. It is equally important that all slave generators exhibit very low jitter. A figure of 5ns or less is typical for timebase stability, while 2ns or less is a reasonable jitter spec.

Where genlocking to an outside source such as network is done on a regular basis, a feature called protected genlock is highly desirable. With protected genlock, the local generator may be locked to an outside source using the independent locking method. Any horizontal timing jumps on the locking signal will not disturb the generator; it will remain frequency locked to the locking signal's subcarrier. If the input signal disappears or reverts to monochrome, the protected genlock logic will instruct the generator to make a smooth switch to the free run mode.

Studio Timing Concepts

In the early days of color, the system designer's attention was focused on the importance of color timing. While timing errors as large as 100 ns could be tolerated in monochrome systems, color systems demanded accuracy two orders of magnitude better. The vectorscope took its place alongside the waveform monitor as an indispensable measurement tool.

As monochrome systems were converted to color, these new timing tolerances had to be addressed. The required degree of precision was not difficult to attain if reasonable care was exercised.

Unfortunately, the importance of the relationship between sync and subcarrier timing, or SC/H phase, were not fully appreciated at the time. It wasn't until the advent of color videotape recording and editing that these issues came to light.

When the NTSC color system was devised, the frequency of the color subcarrier was chosen to be an odd multiple of half the line scanning frequency. This was done so that the subcarrier peaks would occur in alternating positions on adjacent scan lines, thus minimizing the visibility of the subcarrier on monochrome receivers.

The multiplier which was finally chosen was 455/2, or 227.5; this is also the number of cycles of subcarrier per scan line. Since there are 525 lines per frame, there are 227.5 times 525 or 119437.5 cycles of subcarrier per frame. The extra half-cycle of subcarrier means that it takes two frames (or four fields) for the color sequence to repeat.

When a video tape is edited, the machine must ensure that the continuity of this sequence is preserved. If it is not, there will be a noticeable horizontal shift in the picture at the edit point. Many machines incorporate *color framers* to prevent this from occurring. They operate by matching the color framing of the signal recorded on the tape with that of the house reference.

Reliable operation of these color framers is dependent upon the maintenance of consistent SC/H phase throughout the plant. Furthermore, when tapes that were recorded in another facility are edited, the SC/H phase of the off tape signal must match the local plant's SC/H phase. For this reason, an industry wide standard has been established which defines correct (zero) SC/H phase as the coincidence of the 50% point on the leading edge of horizontal sync and the zero crossing of subcarrier (see Figure 5.4-2). Establishing and maintaining correct SC/H phase begins with having a means to measure it. Many modern waveform monitors provide this capability as illustrated in Figure 5.4-9.

Maintaining SC/H phase depends upon careful selection of equipment and periodic measurements of actual performance. Sync generators with low SC/H phase drift are key elements in any SC/H phased plant. It is equally important to avoid system design techniques which might lead to SC/H phase problems (such as the use of regenerative pulse or subcarrier distribution amplifiers). SC/H phase should be checked as routinely as video levels or color phase.

Video Timing Tools

Numerous approaches to plant timing exist, but they all fall into one of two general categories: adjusting the relative phase of the source timing reference or delaying the source video. In general, it is always better to move the timing signal around because even the best video delay systems introduce a certain amount of signal distortion. Obviously, video delay is not a viable solution where a source must be advanced in time with respect to other sources.

Video delay lines are often used when a single source must feed different timing points. This is quite often the case in a system where a source must appear as a direct input to a studio production switcher, but the same source may be selected on a routing switcher which, in turn, feeds the production switcher (see Figure 5.4-10).

In general, it is better to use a delay DA than a passive delay line/video DA combination, because a properly designed delay DA will include circuits which compensate for response errors in the delay elements. Where passive delays are used, they should be of the highest possible quality and the shortest electrical length required to get the job done.

Other devices which can be useful in plant timing are regenerative pulse and subcarrier DA's. They should be applied with caution, however, since they can create system SC/H phasing problems.

Guide to Plant Timing

At first glance, timing a television plant may appear a formidable task. Like anything else, though, the task can be whittled down to size through a systematic method comprised of analysis, synthesis and execution.

Analysis entails the careful examination of the timing requirements in each individual operating area of the plant. This information, along with a clear understanding of the capabilities of the equipment, can be used to synthesize a plan for overall plant timing.

The demands placed on most plants are constantly changing and ever increasing today's special setup may be the standard setup tomorrow. It is important to realize, therefore, that the plan which has been put together so carefully may change. If this happens, it

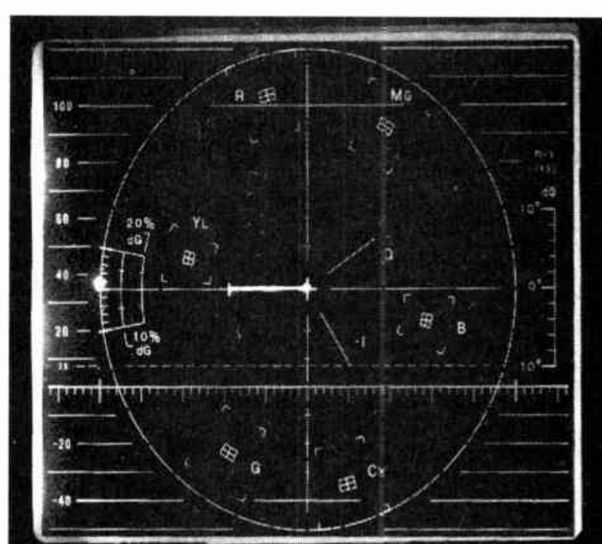


Figure 5.4-9. SC/H phase measurement on the Tektronix model 1750 waveform and vector monitor shows the relation between the burst phase vector and the "sync dot" as either an absolute value or relative to an external reference. Phase may be read out from the graticule or by adjusting and reading calibrated controls. *(Courtesy Tektronix, Inc.)*

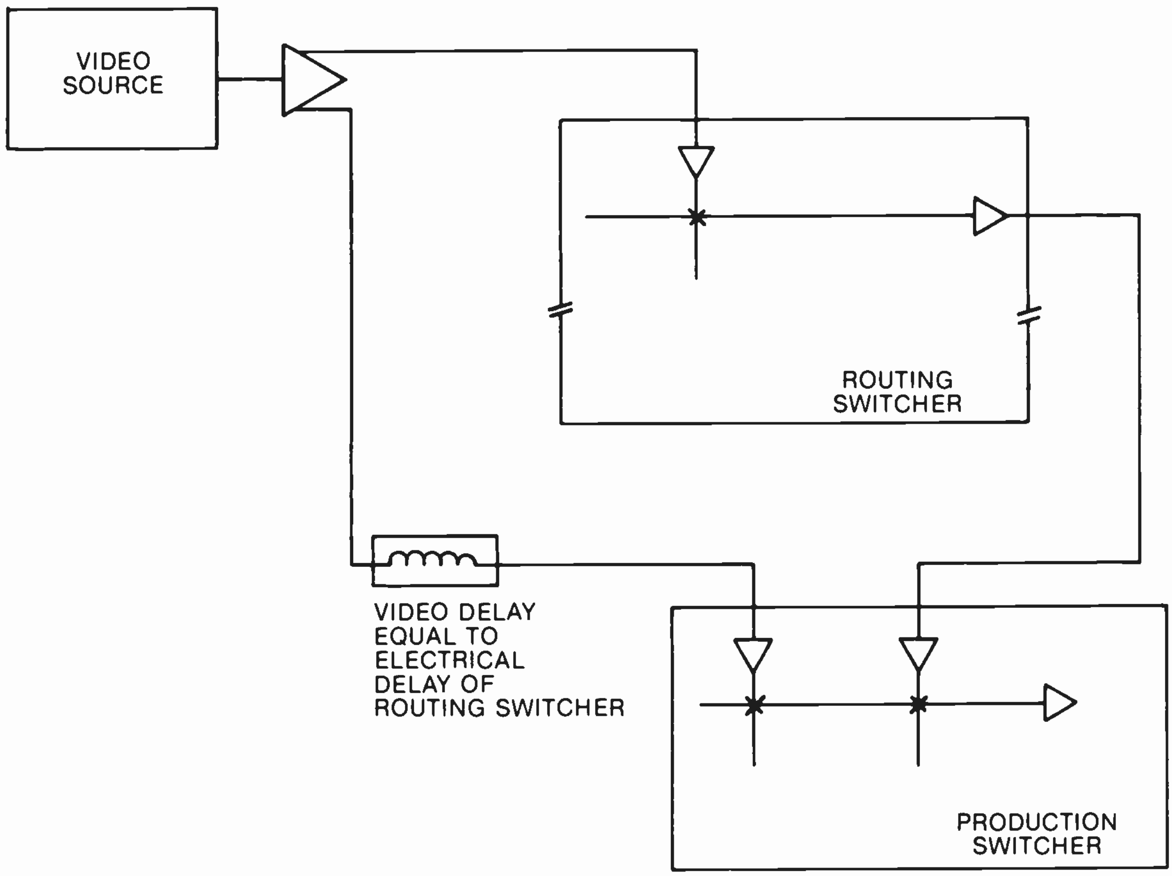


Figure 5.4-10. Timing a single video source to two different timing points.

may be necessary to add or replace equipment or perhaps even alter the system design.

Studio Timing

The studio (or edit suite) is an excellent place to begin a discussion on system timing, since it is the place where proper source-to-source timing relationships is most important. This is due to the fact that the studio is the primary area of the plant where multiple sources are combined. If visible picture jumps and/or color shifts are to be avoided when making mix, wipe or key transitions, all studio sources must be carefully timed.

Timing the inputs to a studio switcher is a very simple, straightforward process:

1. Identify those sources which do not have timing adjustments or whose timing can only be adjusted over a limited range. These will more than likely include the black and background generators in the production switcher itself and may also include test signal generators character generators, and older studio cameras.
2. Using delay DA's or delay lines, match the timing of all of the sources identified in item 1 to the latest in the group of fixed sources. Designate one of these sources as the reference source for setting the timing of all of the adjustable sources. Note that if there is only one fixed source, it should be used as the reference source.
3. Adjust the timing of each of the adjustable sources to match the timing of the reference source.

All timing adjustments should be made by selecting each source, in turn, on the production switcher. An externally referenced waveform monitor and vectorscope connected to the output of the production switcher provide the necessary monitoring capability. Timebase and/or gain magnification should be used on the test instruments to ensure the accuracy of timing adjustments.

The output SC/H phase may change when timing adjustments are made. If the plant is to be SC/H phased, it is imperative that this parameter be checked whenever source timing is adjusted. For equipment that uses separate pulses and subcarrier for timing, care should be taken to maintain a consistent phase relationship between the input sync and subcarrier once the output SC/H phase of the source is established.

If the plant is equipped with multiple studios, the procedures outlined previously should be repeated in each studio. On those sources which feed more than one studio, extreme care should be taken to ensure that cable lengths are made equal. If this is not done, it will be impossible to achieve proper timing in all locations.

Timing in Master Control

For all intents and purposes, master control is just another studio and can therefore be timed using the previously mentioned procedure.

One timing problem which is fairly common in master control is that of studio delay. It generally shows up as a slight timing disturbance when the master control operator switches from a given program source to that same program source through a studio switcher (a common occurrence going into or coming out of a live program such as news). An easy solution to the problem is to resynchronize the output of each studio (see the *Cascading Studios* section of this chapter).

Timing Through an Assignment Switcher

If routing switcher buses are used as input preselectors for a studio switcher, the propagation delay through the routing switcher must be taken into account when timing the plant. This is usually accomplished by setting up both direct and delayed paths for any sources which feed a production switcher both directly and through a routing switcher (see Figure 5.4-10).

In plants where the routing switcher is used for source assignment, it is a good idea to designate the output of the routing switcher as the zero timing point for the plant. If this is done, a bus on the switcher can be designated the plant *quality control point*—one location (with one set of accurately calibrated test instruments) where the signal levels and timing relationships of all sources in the plan can be observed and compared. The generalized timing procedure for studio switchers can be used to time sources into the plant routing switcher.

Cascading Studios

Occasionally, it may be necessary in some plants to cascade studios. In order to preserve overall system timing, the delay through each studio must be taken into account. One approach to keeping everything in time is to time all sources into the first studio *early* by an amount equal to the studio delay. However, if more than two studios are involved and timing into a routing switcher and/or master control are a factor, alteration of source timing may not be practical.

A far better approach is to resynchronize the output of each studio using a frame synchronizer. The timing of each studio can then be set to match that of all other sources in the plant.

A word of caution is in order, however: since nearly a frame of video delay is required to resynchronize a source which is only 100 ns or so *late*, the video from the studio when passed through a frame synchronizer will be out of sync with the audio by approximately 33 ms. While this amount of time difference might not easily be perceived by most individuals, the cumulative effects of several passes probably would be. Therefore, it is a good idea to insert an audio delay line in the program output from the studio.

COMPONENT ANALOG VIDEO (CAV) SYSTEMS

It was once a safe assumption that all video signals throughout a broadcast facility were 1 V p-p standard NTSC signals. However, the widespread use of component analog VTR's in broadcast applications represents a potential challenge to this assumption.

One in. type C and 3/4-in. U-matic VTR's record an NTSC signal on the tape. Half-inch broadcast machines (Betacam, Betacam SP, and MII) record a component analog video (CAV) signal on the tape. These machines have NTSC inputs and outputs and are designed to drop into an NTSC environment, but CAV inputs and outputs are also available.

The chief advantage afforded by this equipment is very high picture quality in a compact, cassette based format. First generation quality comparable to that of 1-in. is possible. The quality of successive generations is good and can be improved considerably if dubbing (and any attendant picture manipulation or processing) is done in the component domain.

In order to extract the most quality from these new tape formats, it is best to avoid repeated conversions to and from the NTSC domain. This can be accomplished quite easily, even in an existing all NTSC plant, by creating CAV *islands*.

CAV camera/recorder units are used extensively in electronic news gathering (ENG) and electronic field production applications (EFP). They are lightweight, relatively low cost, and produce very high quality pictures. Source material recorded in the field is nearly always edited before being used on air. Often, digital effects, captions, and other graphics are added to the finished product. This makes the ENG/EFP edit suite a good model for a CAV island.

The rules of good system design outlined earlier in this chapter apply generally to component systems. There are a few differences, all related to the nature of the component signal itself. NTSC is a one wire system; CAV is a three wire system. All color television signals originate in component form, usually as RGB. The NTSC system was developed to allow these color components to be transmitted over a single transmission channel. And since a single wire is more convenient to deal with than three, NTSC has been the distribution standard as well as the transmission standard. In small islands, the inconvenience of three wires versus one wire is a small price to pay for the enhanced picture quality that CAV offers.

In designing a CAV island, the choice must be made as to which component set the island will be designed around: RGB or Y, R-Y, B-Y. There are pros and cons associated with both, but most system designers lean toward Y, R-Y, B-Y. This is because the Y signal provides a convenient feed for monochrome preview monitors and because the + in. CAV recorders utilize the Y, R-Y, B-Y component set.

In the Y, R-Y, B-Y component set, the Y channel carries the luminance information. The Y channel signal is nominally 1 V p-p and looks very much like a conventional NTSC signal, except that it is monochrome and therefore contains no subcarrier or burst. The other two channels (the color difference channels) contain the color information. They are quite different in appearance from a conventional video signal. First, they do not contain sync. Second, they are bipolar signals with a nominal level of plus/minus 350 mV.

CAV Signal Distribution

When distributing CAV signals, a number of their unique characteristics should be considered. Since they are monochrome signals, overall source timing is not as critical as with NTSC. Therefore, rather than being concerned with timing accuracy to within a nanosecond, the system designer can think in terms of ten or more nanoseconds. Interchannel timing should be maintained as tightly as possible in order to prevent picture impairments due to differences in luma/color timing. Also, the bipolar characteristics of the color difference signals must be preserved throughout the system. Blanking level must be maintained as close to 0 VDC as possible. Since the color difference signals do not contain sync, clamps or dc restorers cannot be used. Instead, it is best to use straight dc coupling. Many DA's and switching systems offer this capability, usually by setting straps or jumpers. Further, interchannel gains should be closely matched and maintained as tightly as possible. Interchannel gain errors or dc shifts will result in color errors in the final picture. These errors can be avoided through careful setup and maintenance of system levels and the selection of distribution and switching equipment with good gain and dc level stability.

Interformat Conversion

Regardless of which component set is selected for the design of the island, some interformat conversion will undoubtedly be required. For example, assume that a CAV edit suite is designed around the Y, R-Y, B-Y component set. Integrating an RGB character generator into such a system will require interformat conversion from RGB to Y, R-Y, B-Y. This is a relatively simple process which can be accomplished using a variety of commercially available products.

In some instances, conversion in the other direction, from Y, R-Y, B-Y to RGB, might also be required. Fortunately, this is also a relatively simple process.

Conversion between the CAV and NTSC environments may also be a consideration. In a CAV edit suite, for example, it may be necessary to import or export NTSC signals. This can be accomplished quite easily by means of the tape medium. But if real time ingoing and outgoing NTSC feeds are anticipated, decoding and encoding equipment will be required. Commercially available NTSC decoders and encoders span a wide price and capability range. In choosing among the alternatives, the system designer must carefully evaluate the level of performance required.

There are basically two types of decoders: *notch filter* and *comb filter*. Notch filter decoders are relatively inexpensive and are fairly good at removing residual subcarrier from the luminance information. The basic characteristics of a notch filter do cause some loss of horizontal resolution in the luminance channel. Comb filter decoders are more expensive but provide excellent suppression of residual subcarrier. Luminance horizontal resolution is not usually affected, but apparent vertical resolution can be impaired, especially in areas of the picture where there are sharp horizontal transitions.

A third type of decoder, the so-called *adaptive comb filter decoder*, combines the best characteristics of both types of filters. Adaptive decoders contain circuitry which dynamically selects the most appropriate filter characteristic, notch or comb, based on picture content. As might be expected, adaptive decoders are relatively expensive.

A similar range of choices exists with respect to NTSC encoders. Inexpensive units employ simple filters, while more expensive units employ more complex filtering techniques. The quality of the encoded picture varies directly with the cost of the box.

Other CAV System Considerations

Unfortunately, there is not a universal standard for CAV. One area of considerable variability is the precise signal levels for the Y, R-Y, B-Y component set. At the time of this writing, there are three standards: SMPTE/EBU, Betacam, and MII. The differences between these standards are relatively small and fall within the adjustment range of all affected equipment within the CAV system.

Timing within a CAV island is relatively straightforward. If the island is to be timed into a large plant, there are certain things which must be taken into account. Chief among these is decoder delay. A simple notch filter decoder will have an electrical length of several hundred nanoseconds. An adaptive decoder may be as long as several horizontal lines. If decoded feeds are to be in time with sources local to the island, the entire timing reference for the island must be delayed (with respect to plant zero time) by an amount equal to the decoder delay. This can be accomplished quite easily through the use of a dedicated sync generator for the island. This generator would be locked to plant sync and adjusted to provide the correct amount of delay.

If the output of a CAV island is to be timed into the plant, the island sync generator could be adjusted to advance the timing within the island by an amount equal to the delay through the encoder placed on the output of the island. If the island is to be fully timed into the plant—on both the input and output side—the island output would have to be re-synchronized using a frame synchronizer.

Test and monitoring in the component domain requires specialized equipment. Conventional NTSC equipment is not suitable. Manufacturers offer a range of equipment specifically tailored to CAV applications.

Figure 5.4-11 depicts a representative CAV edit suite and illustrates some of the concepts covered in the text.

DIGITAL SYSTEMS

Digital video equipment has been around for many years. Electronic character generators, first introduced in 1970, were among the first devices to use digital technology. The first video product to employ digital circuits for signal processing was the timebase corrector, introduced in 1974. The first digital frame synchronizer followed in 1975, and the first digital video effects system in 1976.

This equipment used digital technology for one of two reasons: the superiority of a digital solution over analog (TBC's), or the inherent impracticality of an analog solution (frame synchronizers). All of these early devices were intended to be used in an analog environment and therefore had no digital I/O.

As the variety of commercially available digital video equipment increased, so did the interest in digitally interconnected video systems. The push for digital I/O shifted into high gear in the mid 1980's with the impending availability of the digital videotape recorder (DVTR). In 1987, when the first commercially available digital recorder was introduced, many manufacturers had already incorporated digital I/O into their products.

Digital Video Standards

The majority of digital video equipment available today is equipped with both analog and digital I/O. Digital interconnectability has been made possible by a set of internationally agreed upon standards, many of which have been prompted by the standardization of digital video recording formats.

Digital video equipment exists for both component and composite systems. The first digital video recording format, the so-called D1 format, records a component signal on the tape. Most high quality digital video effects and electronic graphics systems employ component processing. Therefore, if these devices are equipped with digital I/O, it is likely to be component digital I/O. Component signals which have been recorded digitally suffer virtually no degradation from generational losses or encoding/decoding artifacts. This is why component digital is the format of choice for critical graphics work or high end post production.

The high cost of D1 VTR's have made them prohibitively expensive for use in many applications. For example, program and spot playback in broadcast facilities. The need for a more economical digital videotape format prompted the development of two composite digital formats, D2 and D3.

Since then, a number of additional digital video recording formats have been introduced, and all are component formats. They include D5, D6, DCT, Digital Betacam, DVC Pro, and Digital-S.

Interconnection Conventions

Initially, parallel interconnect was the only method available to the digital video system designer. The SMPTE 125M and 244M standards specify a method which uses a separate pair of wires for each data bit and a separate clock. The connector called out in these standards is a 25 pin "D' connector. SMPTE 125M is the standard for component equipment and systems. SMPTE 244M is applicable to composite equipment and systems.

There are numerous drawbacks to the parallel interconnect method. First, the cable is considerably more expensive than coax, both in terms of the cost of the raw cable and the cost of applying connectors to it. Second, the cable and connectors are more bulky than coax. Finally, the distance over which parallel data can be reliably sent is limited to 100 meters maximum.

For these reasons, the vast majority of digital video equipment available today is equipped with serial I/O. SMPTE 259M specifies the mechanical and electrical characteristics of this interface as well as the formatting of the serial data stream. The connector called out in the standard is the familiar BNC. The data rate for serialized composite digital signals (525/60 NTSC standard) is 143 Mbits/sec. The data rate for component digital is 270 Mbits/sec.

Integrating older equipment with parallel I/O into a serial digital system may be accomplished using external conversion devices. Standalone serializers and deserializers are available from a number of different manufacturers.

Standard high quality video cable such as Belden 8281 may be used for serial digital interconnects. Cable manufacturers have begun to introduce cable types that are designed specifically for serial digital video. Belden 1505A is one example of such a cable. Because of the high data rates involved, a low loss double shielded cable is recommended. Runs of up to 300 meters are possible with low loss cable.

Serial digital interconnect cables should be terminated in high quality 75 Ω BNC connectors. For many years, it has been a common practice to use 50 Ω BNC connectors in analog systems. Depending upon its design, a 75 Ω female BNC connector may be damaged when mated with a 50 Ω male BNC. This is due to the larger pin diameter on the 50 Ω connector. Therefore, when existing analog system cabling is used to interconnect serial digital equipment, 50 Ω connectors should be replaced with 75 Ω connectors.

When distributing serial digital signals, plan on using a distribution amplifier. Because of the high data rates, passive loop throughs are not possible on serial digital equipment. Some equipment may have active loop throughs, but this is the exception rather than the rule.

Impairments in Serial Digital Systems

Digital video signals are more robust than analog signals and are therefore less susceptible to hum, noise,

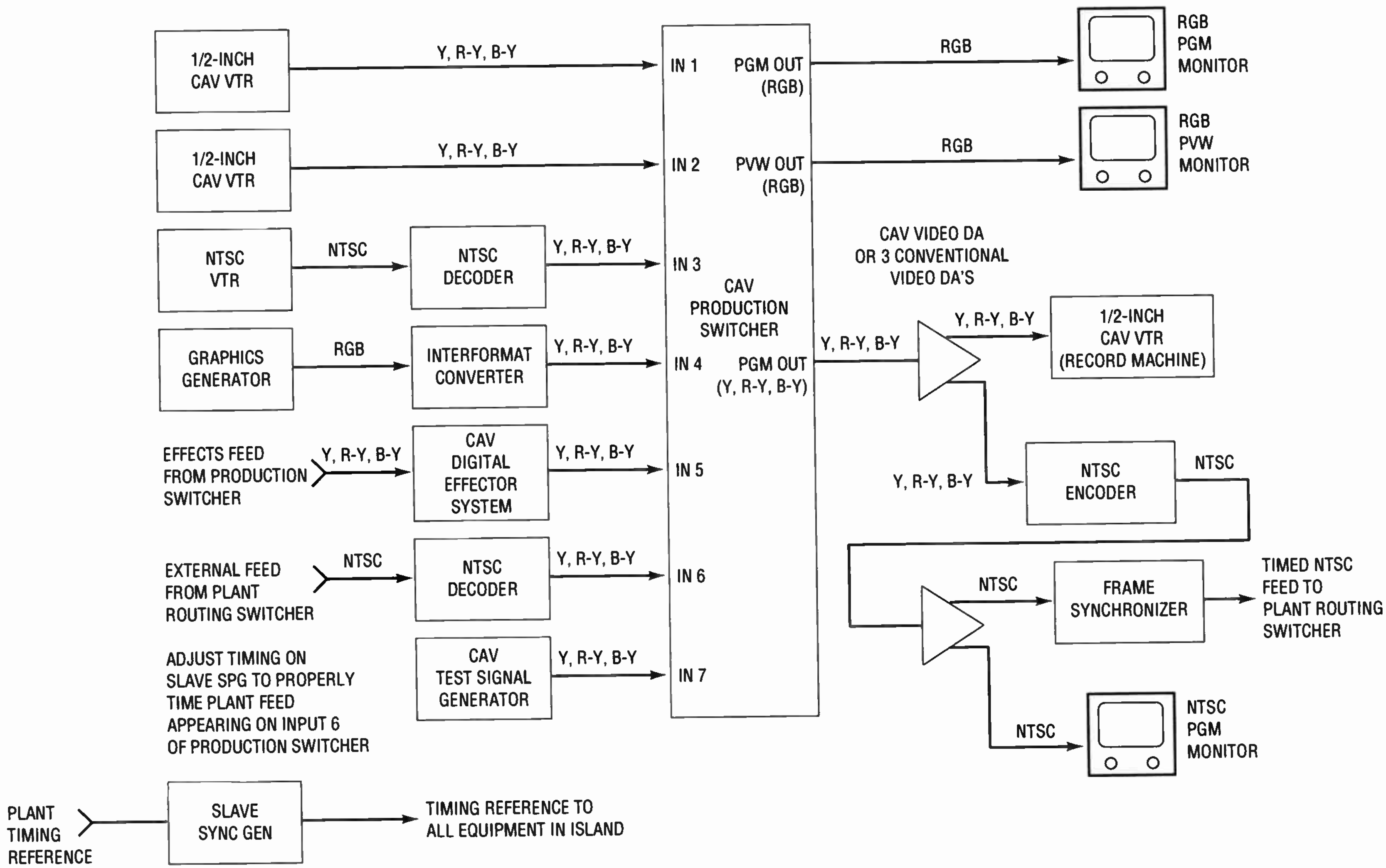


Figure 5.4-11. A typical CAV *island* within an NTSC plant.

crosstalk and other such impairments. Gain variations, frequency response errors and drift are analog problems which disappear in the digital environment.

Cable equalization is still required, but it is of a much different sort than that required in analog systems. Analog cable equalizers must be designed to ensure flat response from nearly dc to 8 or 10 MHz. There is very little low frequency information in a serial digital signal, so the cable equalizer can be designed to compensate primarily for high frequency losses. Flatness of frequency response is not important. And since the data stream can be (and usually is) regenerated, gain adjustments are not necessary. In practice, in a properly designed digital system, all cable equalization and gain adjustments are automatic.

One problem which may occur in serial digital systems which is unique to the digital domain is *clock jitter*. The serial digital signal contains an embedded clock signal which must be extracted to recover the data. As the signal passes through various devices in the system, the data stream may be subjected to repeated regeneration. This has the potential of adding significant jitter to the signal, making clock recovery difficult if not impossible. The solution to this problem is to deserialize the signal then reserialize the parallel data using a stable clock.

Figure 5.4-12 illustrates how jitter affects the serial digital signal. The waveforms show the characteristic eye pattern that is associated with a self clocking serial data stream. The waveform on the left is of a clean, jitter free signal. The eye pattern is well defined. The waveform on the right is of a signal with a moderate amount of jitter. The eye pattern is less defined and the horizontal opening in the eye pattern is closed.

Serial Digital Distribution Amplifiers

A variety of different types of distribution amplifiers are available for serial digital systems. Even the simplest models, designed primarily for use as fan-out amplifiers, have automatic cable equalization. This is because all of the commercially available serial digital receiver IC's have this function built in.

Simple fan-out DA's should be used in applications where reclocking is not required or desired, that is, where the input signal is relatively jitter free. Typical applications would include output fan-out on a VTR, character generator, test signal generator or other source device.

Reclocking DA's should be used in applications where the input signal has jitter in excess of the maximum allowed by SMPTE 259M, 500 ps peak-to-peak. Typically, this would be the case in portions of the signal path that are farthest downstream from the signal source.

There is a phenomenon in digital systems called the *cliff effect*. The term refers to the fact that digital signals do not gradually degrade the way analog signals do. Instead, they suddenly and abruptly fall off the edge of the cliff. One way of monitoring the integrity of a digital signal and to avoid the cliff effect is to measure the bit error rate at various points along the signal path. *Error detection and handling* (EDH) is a method for monitoring and reporting errors in digital systems. It is documented in detail in SMPTE RP-165.

EDH uses hardware to calculate a checksum which is inserted in the auxiliary data space of the serial digital signal. When an EDH coded bit stream traverses a signal path, EDH-enabled hardware can be used to monitor the signal for transmission errors and flag them when they are detected. Serial digital DA's with EDH capability are available from several manufacturers.

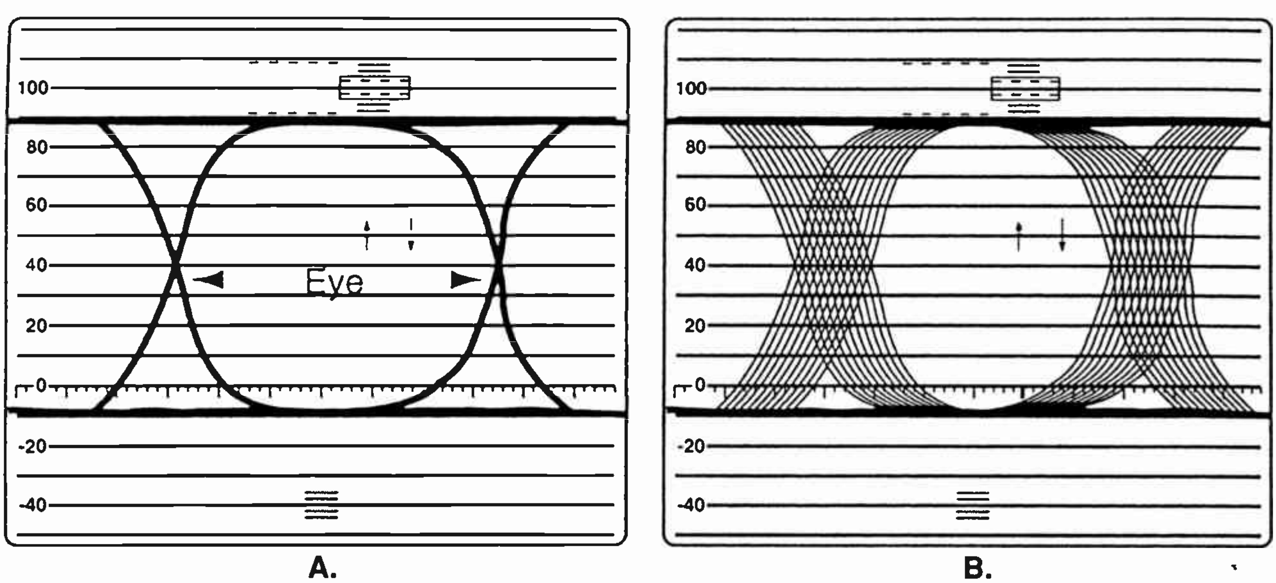


Figure 5.4-12. Eye pattern displays of a serial digital signal, without jitter (left) and with jitter (right). *(Courtesy of the Grass Valley Group)*

Another variety of special purpose DA's of interest to the systems designer is the serial digital DA with analog monitor outputs. Many of the larger color monitors used for program or preset monitors in a control room are available with serial digital inputs. However, most of the smaller monochrome monitors used for source preview monitors have only analog inputs. Using a separate digital to analog converter is one solution to this problem, but it is an expensive one. A far more economical alternative is to use the aforementioned specialty DA whenever an analog monitor feed is required.

Serial Digital Switching Systems

Serial digital video is nothing more than very high speed data. Routing switchers designed to handle serial digital video employ technologies and design philosophies that are radically different from those employed in analog switchers. Switching elements are typically 16x16 (or larger) high speed monolithic crosspoint arrays. Multibus routing switchers are therefore nearly always matrix oriented designs.

Serial digital designs are more compact than comparable analog systems. This is partly due to the density of the digital crosspoint technology, but is also due to the fact that crosstalk in a digital matrix is not as serious a design consideration as it is in an analog matrix. Therefore, digital systems can be packaged more densely.

Important factors to consider when selecting a serial digital routing switcher include: maximum data rate required, discrete data rates supported, expandability, compatibility with other types of matrices (digital audio, analog video, and analog audio) and control system capability.

In an all 4fsc facility, the maximum data rate required is 143 Mbits/sec. However, since 4fsc equipment is not likely to play a significant role in the future of broadcast television, new installations should be built to accommodate the higher bit rates required for component digital signals: at least 270 Mbits/sec, and preferably 360 Mbits/sec.

Many digital routing switchers contain reclocking circuits which are bit rate dependent. Therefore, it important to ensure that any new system can accommodate all standard bit rates up to the maximum rate supported.

The expandability of a routing system should always be given careful consideration. In addition to knowing what the maximum matrix size is, system designers must understand how the system is architected and what will be needed to expand it. In many systems, input fan-out amplifiers, secondary switching matrices, and considerable cabling are needed when expanding beyond a certain size. It is best to have knowledge of these requirements when the initial installation is done, so that they can be planned in advance.

Another important factor to consider is compatibility with other types of matrices. In a transitional, hybrid analog/digital facility, it is often a requirement that the central routing system be capable of handling analog video and audio as well as digital video and audio. Not all matrices need be purchased from the same manufacturer, since many companies provide protocol translators to interface their systems with those from other manufacturers.

The possibility exists in a serial digital routing system to route multiple channels of digital audio along with the video. This technique, commonly referred to as *embedded audio*, is described in detail in SMPTE 272M. Basically, digital audio is formatted into the serial digital video stream during the vertical blanking interval. Embedded audio can provide significant cost savings over separate audio, provided there are enough sources and destinations in the facility which have embedders and disembedders built in. If external devices have to be used to perform these functions, their cost can easily offset any savings in the routing and distribution system.

Timing Considerations

Precise timing is a necessity in analog video systems, particularly where multiple signals arrive at a single point where they will be combined—for example at the input of a production switcher. This is typically not a requirement in digital systems. Most digital switchers and video effects devices have an input timing window with in which all signals are expected to fall. Manual, and in some cases automatic, adjustment is provided to precisely time the signal for the internal processes which the device performs.

Digital signal processing involves longer time delays than analog. As an example, a large analog production switcher has an input to output delay of 500-600 ns. In comparison, its digital equivalent has a delay of a line (63.5 μsec) or more. In order to deal with delays of this magnitude, the systems designer will need correspondingly longer delay lines.

Digital video effects devices typically exhibit throughput delays of either a field or a frame. Delays of this magnitude can create lip sync problems with companion audio, not unlike that which broadcasters experienced with the advent of the frame synchronizer. The system designer should take care to ensure that companion audio signals be delayed to match any video signals which are subjected to very long processing delays. Fortunately, delays of several lines up to a frame or more can be implemented cost effectively in the digital domain.

Generally speaking, plant timing in a digital facility (or in a hybrid analog/digital facility) can be approached in the same manner as described earlier in this chapter for an analog facility. Key points to remember:

1. Establish correct source timing at the zero timing point of each studio or island. This is usually the program bus of a production switcher.
2. Establish a zero timing point for the whole facility. This is usually a reference test bus on the master routing switcher.
3. Time the outputs of all studios and islands into the master routing switcher. This usually involves adjusting the phase of the slave sync generators in the studios/islands.

4. Time studio or island sources that appear as direct inputs on the master routing switcher. This usually involves adjusting delay lines which compensate for the delay through the studio/island.
5. Use frame synchronizers to time external non-synchronous sources and studio reentries.
6. Insert lumped delays in any audio circuits as needed to maintain proper lip sync.

ADVANCED TELEVISION

In 1982, the Advanced Television Systems Committee (ATSC) was formed to develop standards for an advanced television system for deployment in the United States. At the time, analog HDTV was being tested on an experimental basis in Japan, and Europe was engaged in the development of its own analog HDTV system. See Chapter 1.11, *DTV Standard*, for more history of the process.

The U.S. DTV Standard

In December of 1996 the FCC adopted essentially all the ATSC standards for digital television. The new standard supports both high definition television (HDTV) and standard definition television (SDTV) in a variety of different scanning formats. The standard also supports multichannel surround sound and data broadcasting. At the time of this writing, there is considerable debate about how broadcasters will choose to program this new service. Two basic options exist: a single HDTV program or multiple SDTV programs for each DTV channel. The most likely scenario is that there will be a combination of both.

DTV System Issues

Uncompressed high definition video requires approximately 1.5 Gbits/sec of bandwidth. Networks and stations that opt to produce and distribute HDTV programs will be faced with many technical challenges. Even though several manufacturers have built or plan to build digital HDTV cameras and VTR's, very little switching and distribution equipment exists outside the laboratory.

The situation is entirely different with respect to SDTV. Existing 270 or 360 Mbit/sec switching and distribution equipment will easily handle digital SDTV. This same infrastructure could be used to distribute compressed HDTV signals within a facility. A compression ratio of approximately 4 to 1 would provide sufficient bit rate reduction while minimizing the effects of cascaded compression codecs.

The transmission bit rate specified in the DTV standard is approximately 19 Mbits/sec. To fit a single HDTV or multiple SDTV programs into this bandwidth requires a high degree of compression, approximately 75 to 1. It will not be possible to decompress such a signal and recompress it without serious picture degradation. Therefore, new techniques will be required to perform simple operations such as video mixing and keying when dealing with highly compressed bitstreams.

Switching compressed MPEG-2 bitstreams is also a nontrivial task. The MPEG-2 compression scheme utilizes I, P, and B frames to achieve a high degrees of bit rate reduction. I frames are complete descriptions of a scene, but P and B frames contain only changes since the last I frame. The sequence of I, P and B frames generated by an MPEG-2 encoder changes dynamically with scene content. Therefore, it is not possible to perform a deterministic switch between two MPEG streams. The two streams must be spliced according to a complex set of rules regarding the frame sequences of both.

Solutions will be developed for many of these system problems as the DTV build out progresses.

BIBLIOGRAPHY

Benson, K. Blair, and Whitaker, Jerry, ed., *The Television Engineering Handbook, Revised Edition*, McGraw-Hill, New York, 1992.

Electronic Industries Association., *EIA Standard RS-170, Electrical Performance Standards—Monochrome Television Studio Facilities*, EIA, Washington, D.C, 1957.

Electronic Industries Association., *EIA Standard RS-250B, Electrical Performance Standards for Television Relay Facilities*, EIA, Washington, D.C, 1976.

The Grass Valley Group (GVG)., Designing Digital Systems, (GVG) Grass Valley, CA, 1994.

The Grass Valley Group., NTSC Studio Timing: Principles and Applications, (GVG), Grass Valley, CA, 1986.

The Society of Motion Picture and Television Engineers., (SMPTE) *SMPTE Standard 125M-1995, Component Video Signal 4:2:2 - Bit-Parallel Digital Interface*, White Plains, NY, 1995.

The Society of Motion Picture and Television Engineers., *SMPTE Standard 244M-1995, System M/NTSC Composite Video Signals - Bit-Parallel Digital Interface, White Plains, NY, 1995.*

The Society of Motion Picture and Television Engineers., SMPTE Standard 259M-1993, 10-Bit 4:2:2 Component and 4fsc NTSC Composite Digital Signals—Serial Digital Interface, White Plains, NY, 1993.

The Society of Motion Picture and Television Engineers., *SMPTE Standard 272M-1994, Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space*. White Plains, NY, 1994.

The Society of Motion Picture and Television Engineers., *SMPTE Recommended Practice RP 165-1994, Error Detection Checkwords and Status Flags for Use in Bit-Serial Digital Interfaces for Television*, White Plains, NY, 1994.

Watkinson, John., *The Art of Digital Video, Second Edition*, Focal Press London, England, 1994.

5.5
MAGNETIC AND OPTICAL RECORDING MEDIA

JOE GREGA
QUANTEGY, INC., PEACHTREE CITY, GA

INTRODUCTION

Recording media has been in existence since magnetic wire recording was invented by Valdemar Poulsen in 1898. The idea of recording on media has resulted in continual evolution of magnetic tape and the introduction of magnetooptical and optical recording media and systems. Historically, data storage requirements have doubled every year and this trend is expected to continue. The next step in the evolution of recording media may be a few years away; it may be charged particle technology using specifically focused ion beams, blue laser optical tape or some other type of new technology.

This chapter focuses on the primary or most common formats of magnetic or optical recording media. The technology or principles of the media types reviewed will cover the majority of the recording media formats that are not reviewed. The most significant trend affecting the entire recording media business has been the storage of more information in less space (packing density). This trend is expected to continue regardless of the media used. Recording technology companies are on a never satisfied quest to increase storage densities, in smaller form factors, and to increase read, write and retrieval speeds.

When selecting a media for an application there are many options and many criteria to consider. Table 5.5-1 provides a basic comparison of a several different types of media.

Brief History of Magnetic Recording

The recording systems of today are based upon the pioneering work of a number of individuals and companies. Some of the more prominent include the following:

- 1888 Principle of magnetic recording discovered by Oberlin Smith
- 1898 Magnetic wire recording invented by Valdemar Poulsen
- 1912 Lee DeForest invented the vacuum tube
- 1927 ac bias discovered
- 1931 Blattnerphone steel band recorder used by the BBC
- 1935 I.G. Farben (BASF) makes tape with iron oxide powder on a plastic film AEG-Telefunken develops the Magnetophone tape recorder
- 1946 The German Magnetophone is demonstrated in the U.S. by Jack Mullin of AMPEX
- 1948 First U.S. audio tape recorder is introduced by AMPEX model 200
  First U.S. magnetic tape is introduced by 3M and Audio Devices
- 1951 Stereo tape recorder, Magnecord (1250) introduced
- 1956 November 30, first broadcast of a videotaped recording, Douglas Edwards and the News.
  The computer hard disc is introduced
- 1958 Color videotape recorder is introduced
- 1963 Philips introduces the audio cassette
- 1975 SONY Betamax is the first successful consumer videotape recorder
- 1981 Camcorder is introduced
- 1984 The audio CD is introduced
- 1986 DAT audio format introduced
- 1992 Audio Data mini disc format is introduced
- 1994 Recordable CD format is introduced
- 1996 Digital versitile disc is introduced

OPTICAL RECORDING MEDIA

The use of optical media has become more widespread since the consumer introduction of the audio CD in 1984. While magnetic tape is still the most popular or commonly used form of removable recording media, its ability to increase packing density appears more limited than optical media and its future, therefore, is limited.

Optical recording media reads and writes information by different means. Reflectivity and Keer effect are the basics underlying principles of optical media. ROM type media is read as the pickup reads the intensity of the reflected laser beam as it is changed by the pit and land surfaces. The signal received from the disc depends on both changes in reflectivity and the polarization angle of the reflected light (Kerr effect). Chemical and physical changes in the dye layer result in changes in reflectivity contrast.

Optical Recording System Description

An optical media system is composed of the following basic components:

- Media drive mechanism—powers the laser, supplies the error correction code, controls the optical head and disc media motor. It also consists of the electron-

Table 5.5-1
Comparison of Various Storage Media

	Phase Change Dual (PD)	Magneto Optical	Removable Hard Disc	Mini Disc Data	Zip	Floptical
Capacity (MBs)	650	128/230	105/270	140	100/25	25/7
Media Size (inch)	4.7	3.5	3.5	2.5	3.7	3.5
Transfer rate	518–1.141 KB/s	1.15–1.84 MB/s	1.40–2.34 MB/s	150 KB/s	12.5 MB/s	0.61–1.6 MB/s
Average seek time	165 micro seconds	27 micro seconds	13.5 micro seconds	300 micro seconds	29 micro seconds	68 micro seconds
30 MB read speed	47 seconds	30 seconds	35 seconds	204 seconds	52 seconds	*
30 MB write speed	71 seconds	68 seconds	50 seconds	210 seconds	53 seconds	*
Write method	Optical Mod.	Optical Mod.	Magneto Field Mod.	Magneto Field Mod.	Magneto Field Mod.	Magneto Field Mod.
Read method	Optical head	Optical head	Magnetic head	Optical head	Magnetic head	Magnetic head
Recording method	Phase change	Magneto Optical	Magnetic	Magneto Optical	Magnetic	Magnetic

ics that amplifies the signal and the performs signal processing.

- Optical (send/receive) head consists of the laser light source, an optical system to focus the laser and a detector to convert the returning light to an electronic signal. Figure 5.5-1 shows the basic components of the optical head.
- Optical media. There are three basic types of optical media, CD-ROM, WORM and erasable.
- Computer system and interface—provides data from numerous sources and controls basic computer functions.
- The amount of data that can be stored on a given disc depends on the computer system and optical drive system.

Optical Recording Media

It is necessary to discuss optical media by a number of defining terms, such as form factor, functionality, and capacity; for example a 130 mm 2.6 GB magneto-optical disc. To more completely describe an optical disc, the following five items should be specified:

- **Form factor.** Describes the physical dimensions of the disc itself and does not include the cartridge dimensions. Common sizes of discs are 3.5 in. (90 mm), 4.72 in. (120 mm), and 5.25 in. (130 mm).
- **Functionality of disc.** ROM = read only—cannot write or erase. WORM = write once read many—can write once and read but not erase. Rewritable = can write, read, and erase many times.
- **Read/write methodology.** Embossed, organic dye, ablative, phase change, magnetooptical.

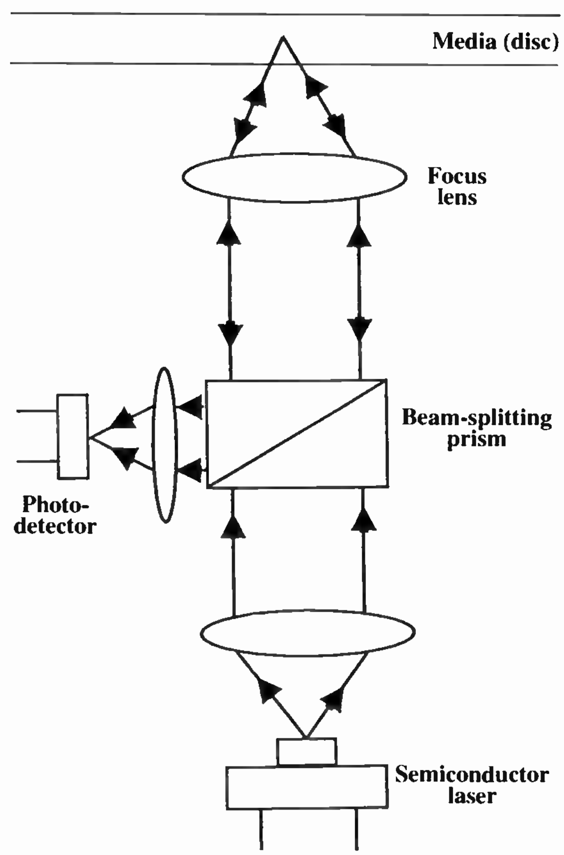


Figure 5.5-1. The basic components of the optical head.

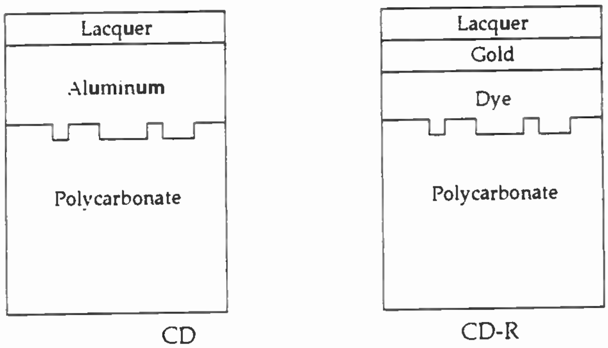


Figure 5.5-2. The basic differences between a typical CD-ROM disc and a CD-R (recordable) WORM disc.

- **Format.** Usually defined at three levels: 1) low level allows the drive to recognize the disc through physical sector layout; 2) defect management and embossed track and sectors; 3) high level formatting provides application-specific formatting.
- **Capacity.** Specified by total storage capacity and sector size, for example, 2.6 GB at 1024 bytes per sector for data application discs, and by recording time in minutes for audio and video discs.

Optical Disc Functionality

ROM devices cannot be rewritten or erased. Laser disc, prerecorded audio CD, computer software CD, prerecorded Mini Disc, and prerecorded DVD discs are examples of this type of media. ROM media consists of a series of pits of varying length that are formed as the polycarbonate substrate disc is molded. Once a CD-ROM disc has been molded the data cannot be altered. A thin reflective film is coated onto the pit surface of the substrate. When light is focused on the pit, it is scattered by edges of the pit. Light reflected from the area between (the land) the pits are scattered very little. The difference between the non-scattered and scattered light is sufficient to be read by the detector. Figure 5.5-2 shows the basic differences between a typical CD-ROM disc and a CD-R (recordable) WORM disc.

WORM discs are best known as the CD-R format, which has become popular for computer storage applications. They are ideal for distributing large amounts of information. The newest technology addition to this group is the DVD-R. WORM discs have a special reflective layer which is melted by the laser to form a pit or hole in the reflective layer. During the forming of the pit or hole the laser is operating in a high power mode. Chemical and physical changes in the dye layer result in changes in reflectivity contrast. The reflectivity difference between the hole or pit and the reflective surface is read by the detector. Figure 5.5-3 illustrates the read and write technology of an Ablative WORM disc.

WREM (write-read-erase memory) is also called erasable or rewritable. Data can be written, erased, rewritten and erased as many times as desired. Magnetooptical media is one form of WREM media that has become popular in recording systems because it can be erased and rewritten over one million times. Examples of rewritable discs are 3.5″ MO, 5.25″ MO, Mini Disc Data, and DVD-RW discs.

Rewritable optical discs can be written on and erased just like floppy discs. Rewritable MO discs are ideal for a variety of uses, ranging from personal computing to video applications such as short-form segment storage.

Magnetooptical Media

The MO disc has a number of layers stamped into its surface during manufacture. In MO technology, magnetic and optical principles work together to enable hundreds of megabytes of data to be stored over many decades without data loss. The Magnetooptical recording process is based on two principles: 1) a magnetic materials coercivity (its resistance to changing magnetic orientation) drops when heated, and 2) laser light polarization is rotated when exposed to a mag-

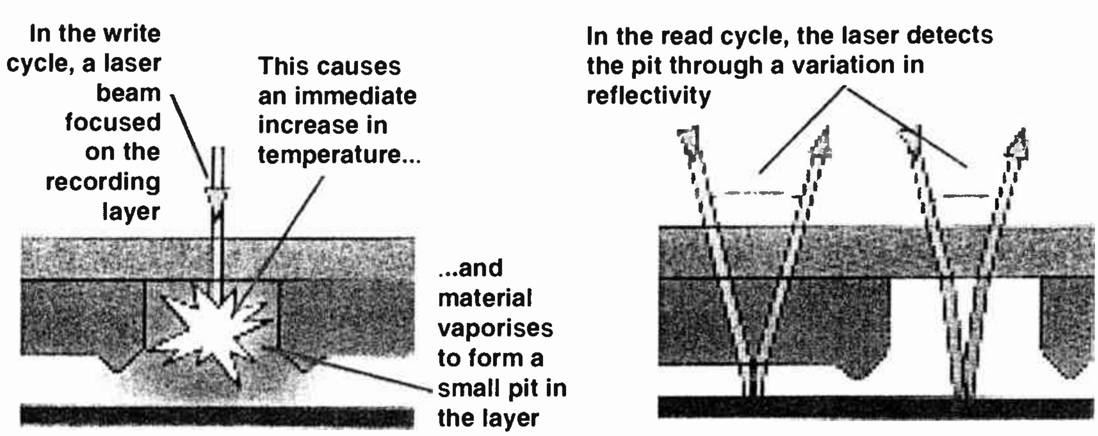


Figure 5.5-3. Read and write technology of an Ablative WORM disc.

netic field. The signal received from the MO disc depends on both changes in reflectivity and the polar Kerr rotation. High data storage is achieved because each data spot is identified by the narrow wavelength of laser beam. At the focus of the beam the local temperature of each spot increases to 200° C, and its coercivity drops from 8,000 Oe to less than 200 Oe. At such a reduced coercivity level, it requires the influence of only a low intensity magnetic field to alter the magnetic polarity of each spot. The information to be stored on the disc is coded by the sequence of north and south polarities, each representing a binary state of "0" or "1."

Figure 5.5-4 shows the basic differences between ROM, WREM, and MO discs. During MO recording, data is written onto the disc when a laser beam is focused to a spot (1 μ) on the spinning disc. The laser heats the magnetic material to the Curie point temperature, the threshold at which its coercivity drops. A magnetic head on the opposite side of the disc is driven by the digital signal to be recorded. The spot heated by the laser takes on the magnetic orientation given by the head. The laser beam in MO does not encode the magnetic material, it only heats the medium to make it more easily magnetized. A switchable electromagnet is used to record the data as either "1s" or "0s." In conventional MO drives, the laser is pulsed while the magnetic bias field, which determines the magnetization of the recording layer, is held constant (see Figure 5.5-5).

During playback, a lower power laser beam is reflected from the spinning disc where it is exposed to the magnetic fields on the recorded disc. These magnetic fields rotate the laser beam's polarization (Keer effect). The reflected beam can be of two alternating polarizations as a result of the north-south or south-north recorded magnetic fields. A polarizing beam splitter directs the two polarizations to photodetectors.

Separate erase and verification passes are usually required before re-recording can be done on a previously recorded MO. Technology to allow direct overwrite has been developed using a few different meth-

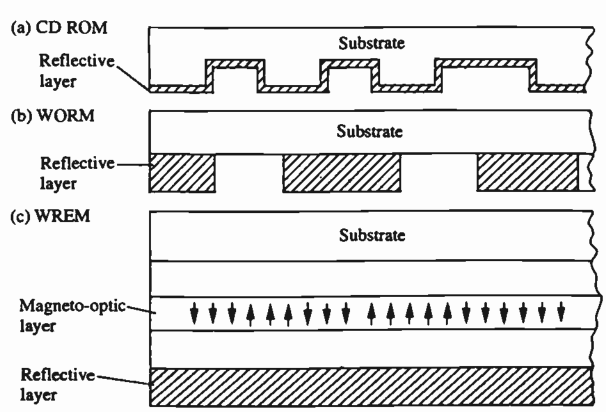


Figure 5.5-4. The differences between ROM, WREM, and MO discs.

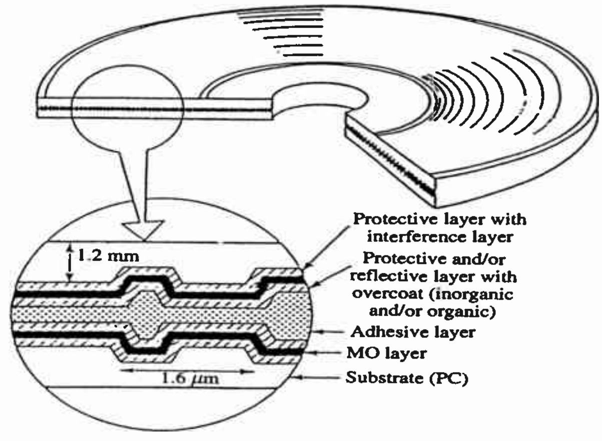


Figure 5.5-5. The conventional MO drive.

ods. MO discs are usually formatted for three distinct purposes. Embossed track and sector formatting and defect management area (DMA)/bad sector formatting are performed at the factory before the disc is shipped. System specific formatting is performed on the actual hardware to be used. Storage capacity of optical media is a function of track pitch, laser beam wavelength, objective lens numerical aperture, data encoding technique, data storage linear density and data storage user area.

There are two types of MO disc compatibility: drive compatibility and logical format compatibility. As long as the device driver and formatting software are recommended by the drive manufacturer there should be no compatibility issue. Optical jukeboxes (or autochangers) utilize a robotic device to insert and remove the optical media from an optical device. Optical jukeboxes are typically used with MO drives and provide as much as 79 Gb of near-line storage per unit. These devices are finding acceptance in most of the standard MO application environments.

Mini Disc

The Mini Disc (MD) is a 2.5 in. optical disc that stores 74 minutes of digital audio. It has a frequency response of 5 Hz to 20 kHz, a dynamic range of 105 dB, and a sampling rate of 44.1 kHz. Mini Disc is not CD compatible as it employs data compression. The MD format uses two types of media: ROM media for prerecorded software and audio and magnetooptical disc for recordable blank discs. The magnetooptical drive is similar to other drives already in use with the added capability of direct overwriting, previously recorded data can be erased and new data written at the same time. To ensure compatibility between discs and players, recordable discs are manufactured with grooves. These grooves guides the writing and reading laser within a spiral track that is 1.1 microns wide. These grooves contain a wobble to create addresses in 13.3 ms intervals (see Figure 5.5-6).

The MD was introduced in 1982 and was targeted at the analog audiocassette and the digital compact

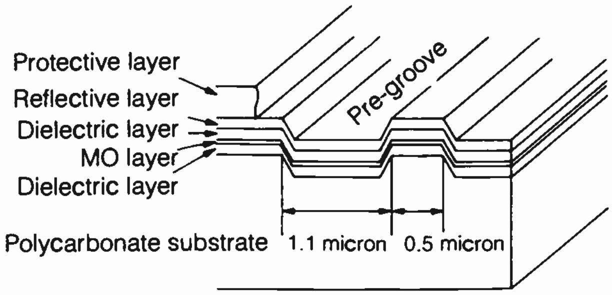


Figure 5.5-6. The Mini Disc.

cassette (DCC). The MD is a consumer product that has the potential of completely refocusing retail music business. It offers the best of features of cassettes and CDs—high sound quality, random access, durability, portability, convenience, shock resistance, and recordability. Professional 4 track systems offer 16 bit audio, MIDI time-code and 37 minutes of recording time of four track mode.

The Mini Disc is also available in a data version for use with the newly developed MD Data Drive format, appropriately called MD Data Disc—with a capacity of 140 MB, a transfer rate of 1.2 Mbps and an average seek time of 300 ms. The MD Data Disc is specifically formatted for data recording and cannot be used for audio recording.

Digital Versatile Disc (DVD)

The Digital Versatile Disc is the latest entrant into recording disc technology. This format actually represents a family of discs using the DVD format and performing different functions.

A DVD-ROM disc is constructed in the same basic way as any ROM disc. The technology of the head and disc has improved and thereby increased the capacity of the data that is stored on the disc. The DVD-ROM format has a capacity of 4.7 GB per side. The DVD-ROM disc is being used for video and software distribution and an audio DVD will soon be available. DVD-R or DVD-RW (rewritable) writers and discs are beginning to appear in the marketplace.

The DVD-R disc is a write once media and depending on the type of data recorded may be capable of being played back on DVD-ROM players. In its initial version, the DVD-R disc will be capable of recording 3.95 GB per side, or six times the capacity of a CD-R disc. Data can be written at 11.08 Mbps, or nine times a CD-R's 1X speed. The write technology employed is similar to that used by CD-R except that the data can be written at a higher speed and the recording density is much higher. To achieve the six times increase in density two components of the recording hardware were changed. The wavelength of the laser was changed to red laser with a wavelength of 635 nm. In the case of CD-R the laser is infrared with a wavelength of 780 nm. The numerical aperture of the lens was improved to 0.5 from 0.6. These changes allow a DVD-R disc to record marks as small as 0.44 um as compared to 0.834 um for CD-R. The recording action happens by momentary exposure of the recording layer to a high power laser (approx. 10 mw) beam that is tightly focused onto the surface. The heated dye polymer layer and is permanently altered so that microscopic marks are formed in the pregroove. The light intensity of the recording laser is tuned to a wavelength of light that is not effected by exposure to ambient light or playback lasers. Playback occurs when a low power laser, of approximately the same wavelength, is focused onto the disc. Figure 5.5-7 shows the basic footprint of the DVD-R.

MAGNETIC RECORDING TAPE

Magnetic tape has been used for over 50 years as a recording media. Since its introduction magnetic tape has evolved in many aspects and diverged into many different types of tape. Particulate iron oxide coating has been the backbone of the magnetic recording industry for most of these 50 years. As recording technology has evolved the magnetic tape has had to evolve to increase its packing density. The ability of magnetic tape to increase in information packing density has been possible through improvements in magnetic particle and recording head technology. Smaller sized magnetic particles and more consistently sized (narrow distribution) particles in combination with improved formulation chemistry and processing procedures have aided the ease of manufacture of these smaller particles. Initial development to improve packing density was focused on increasing the magnetic coercivity of the oxide itself, by increasing the amount of magnetic oxide in the coating and by changing the type of magnetic material. The magnetic portion of a typical particle is approximately 50 volume percent. The actual magnetic part of a typical magnetic coating may be as low as 31 volume percent. Pure iron (metal particle) began to appear in the industry in the mid-eighties. These particles are pyrophoric by nature (they spontaneously burn on exposure to air) so the outermost layer are treated (coated) by the manufacturer. The resultant particle is 74 volume percent magnetic in a coating that contains 45 volume percent magnetic material.

The orientation of magnetic particles greatly determines the recording properties of magnetic tape. Developments in coating and processing technology to improve the orientation have contributed to increases in packing density. The direction of magnetization in a standard iron oxide particle is parallel to its long axis. The more uniform the orientation, the more discrete the transitions are on playback, giving better signal-to-noise properties. As coercivity continues to increase, particle size continues to fall. Today's metal particles are considerably smaller and have much less surface area (a major factor in determining the amount of binder used in a formulation). Smaller particles obviously mean that a greater number of individual parti-

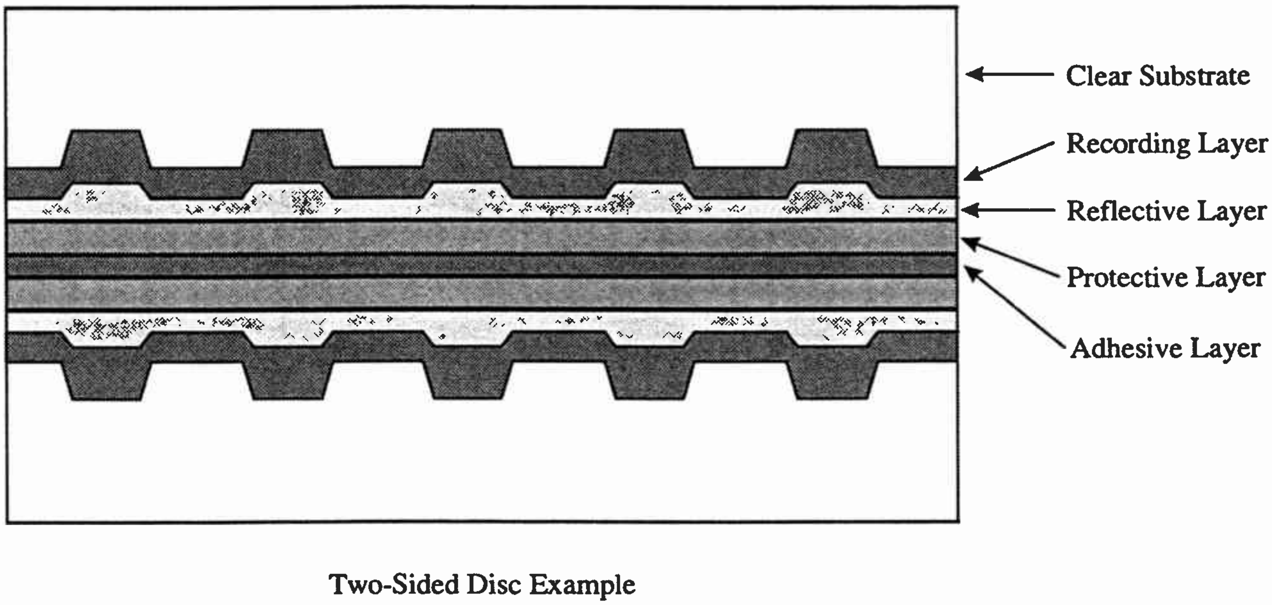


Figure 5.5-7. The basic footprint of the DVD-R.

cles can be located within a given area. This fact, along with the higher coercivity and retentivity, allows a much greater packing density rate, that is, a greater amount of information to be stored in a given area. More uniform particle shapes and sizes and better dispersant or milling technology has also contributed to packing density increases. Figure 5.5-8 shows the increases achieved in some of the more recently introduced videotape formats.

Recording Format Trends

In the development of newer tape recording formats there are four major trends: digital, component, cassette and compactness. While digital recording has been standard in the data field, it is a major trend in

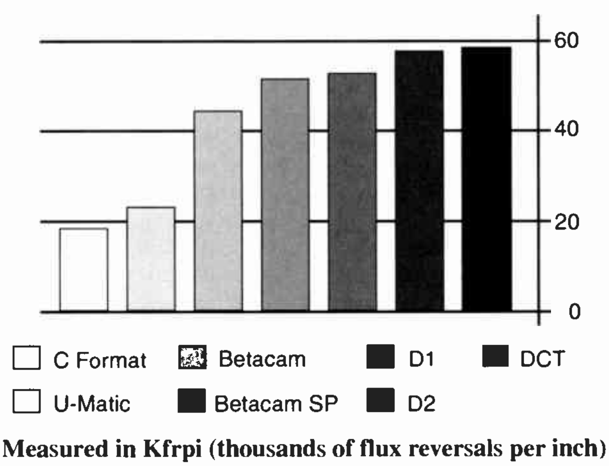


Figure 5.5-8. A comparison of packing density in videotape formats.

the video and audiotape marketplace. The movement is away from the traditional analog recording and toward digital recording formats. Most of the more recent tape based format introductions have been digital recording formats. This is clearly a desirable trend as far as the quality of the audio and video to the marketplace.

Another format trend, specifically in the videotape marketplace, is the movement away from composite recording systems toward component recording. Component recordings are generally of a higher picture quality than composite recordings. Composite signal systems record one signal onto the tape, whereas component signal systems record three—a luminance signal and two chrominance signals. As a result, component systems must record more information onto the tape surface within a given physical space.

The third major format trend is the movement from reels toward cassettes. Most new magnetic tape formats introduced in the last ten years have been cassette based.

The fourth major trend is compactness. Newer formats are becoming progressively smaller, both in terms of the cassette size and the track widths recorded onto tape. As cassettes become increasingly smaller, the complexity of their design keeps increasing, both in the physical number of component parts contained within them, and, the interaction of these component parts. Recent format introductions such as DVCPRO, Digital S, and Betacam SX all continue these trends.

Physical Attributes Trends

Figure 5.5-9 shows the changes in the physical width of videotape formats over time. The earliest format, quadruplex was 2 in. wide, whereas, the newer digital formats have seemingly settled on ½ in. and 19 mm

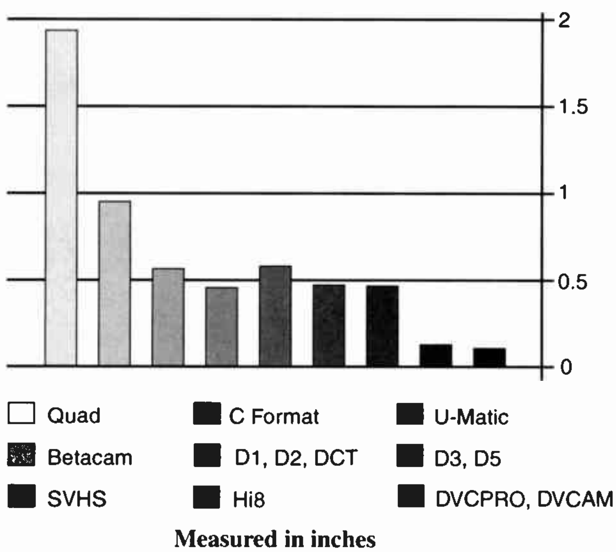


Figure 5.5-9. A comparison of the physical width of various videotape formats.

as the most desirable width. Even narrower widths have been introduced in the professional market.

Figure 5.5-10 shows the changes in the physical thickness of the various formats over time. Newer formats are becoming increasingly thinner. This also represents a series of challenges to the tape manufacturer. With the reduction of the thickness of magnetic coatings, different coating methods have been required for magnetic media manufacturers.

An additional factor that is effecting coating thickness is the trend to digital formats which use very high frequencies in recording. The use of very high frequencies requires only the uppermost portion of the tape surface to be highly magnetic. As the lower portion is unused for recording its purpose is to add some tensile properties that additional thickness brings to the tape. Applying relatively thick coatings in digital is an unnecessary practice and raises the cost of manufacture. Consequently, coating thickness becomes thinner as the trend toward digital formats continues.

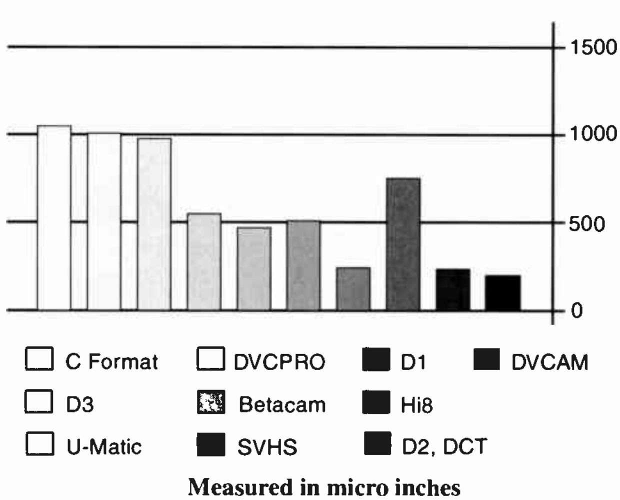


Figure 5.5-10. A comparison of the physical thickness of various videotape formats.

THE MANUFACTURE OF MAGNETIC TAPE

The manufacture of magnetic recording tape is a complex process involving a wide variety of sciences and manufacturing disciplines. The manufacturing process is essentially linear in concept. One step of the process follows its predecessor and cannot be started until its predecessor is complete. The manufacturing process, varies slightly from manufacturer to manufacturer, depending on the physical plant equipment and preferences. Consequently, the description of the process that follows is a lower level, basic or generic process. It provides an overview of the manufacturing process, not an in depth discussion on any or all steps. The following flow diagram represents a basic magnetic tape manufacturing process (see Figure 5.5-11). One element that is not shown is that of time. The entire coating of one batch or mix may take one week, depending on the amount of time required for milling, heat curing and final finished goods assembly.

Formulation Components

Magnetic tape is typically made up of three parts or layers: backcoat, magnetic coating, and base film. The formulation of any magnetic layer contains several components. Each component is chosen for how well it performs its specific function and its synergy within the entire formulation. Although the types of components may differ from one formulation to another, essentially all magnetic recording tape coatings contain most or all of the following components. The final formulation in its wet, coating state has a consistency similar to that of fingernail polish.

Magnetic Media (Pigment)

The type of magnetic material used during the evolution of magnetic recording tape has changed dramatically. Early audiotape products used ferric oxide (FeO_2) with coercivities in the 300–350 Oe range. In search of higher packing density capabilities tape manufacturers have moved toward high energy particles such as gamma ferric oxide, cobalt doped gamma ferric oxide, metal particles, and cobalt nickel for metal evaporative tape. The trend is clearly toward magnetic materials exhibiting high coercivities (1500–2500 Oe) and high retentivity performance. Generally, magnetic particles can be described as acicular in shape (similar to a sewing needle). The smaller the size of the metal particle, the larger the quantity of the particle that can be packed into a given surface area on the tape. This allows much higher data packing densities. The first magnetic recording tape, in 1935, used magnetite (Ferroso-oxide Fe_3O_4), which proved to highly unstable. The magnetic material used has evolved to gamma

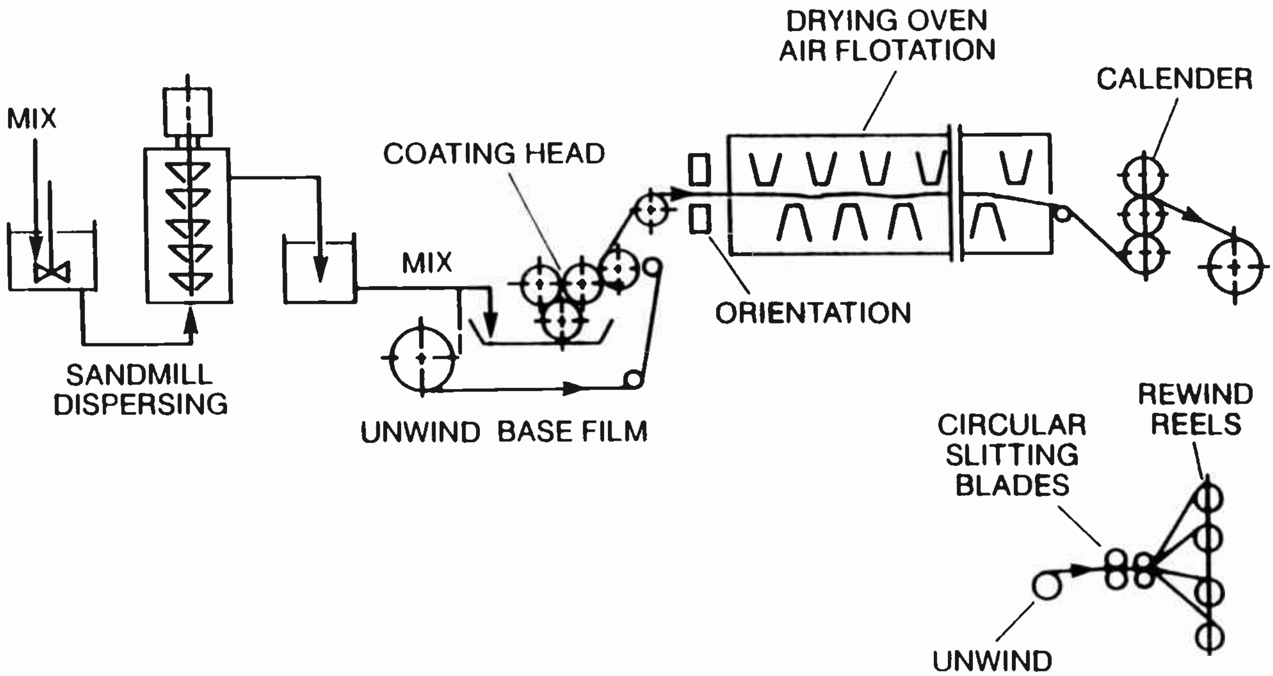


Figure 5.5-11. Schematic of mix preparation, tape coating and processing.

ferric oxide, chromium dioxide, and magnetic particles. Figure 5.5-12 is a representation of the crystal structure of a gamma ferric oxide and metal particle to show the difference in the density of the iron atom versus the oxygen atom. A typical gamma ferric oxide particle is 25 to 35 μ in. in length and about 7 to 12 μ in. in width. For comparison, a cigarette smoke particle is about 25 μ in. in diameter.

Binder

The binders used in the formulation can be likened to paste as they provide the cross linking necessary to keep all of the ingredients together (cohesion) and to ensure the coating does not shed or flake off the base film (adhesion). Binders also add considerably to the stiffness of the finished tape product. Stiffness is required so that the tape wraps the helical scanner and transport assemblies properly and adequate head-to-tape contact is maintained.

The term binder is a generic term covering a wide range of synthetic polymeric material such as polyurethane, elastomers, and vinyl chloride. A recent trend is toward the use of a functionalized binder system. This type of binder system more effectively provides the cross-linking necessary to achieve both thinness and durability as compared with more traditional non-functionalized binder systems. Functionalized binder systems allow the formulation to use less solvent and dispersant, thus simplifying the formulation chemistry.

Plasticizer

The design geometry of modern tape transports results in tape to guide wrap angles of greater than 90°. This is particularly true in most helical scan (head angle across the tape surface) transports. Because of these guiding requirements, some of the forces placed on the tape are in a plane other than vertical. This type of physical requirement places different demands on a tape product than in the past. There is a clear need for the tape to be flexible, yet have relatively high coating modulus. Plasticizers typically perform this function. Without plasticizers, tape would be somewhat brittle and unable to survive the high physical demands and speeds of a modern tape transport. While equipment manufacturers design their transports to place the minimum possible physical requirements on the tape, their priority in design is in the more critical footprint and system performance areas.

Head Cleaning Agent

Typically, head cleaning agents are added to formulations. Usually, in the form of an alumina particle of a specific size. The purpose of these particles is to ensure the heads of the machine remain clean. In addition, these cleaning agents keep the remaining portions of the tape path free from dirt and contamination.

There are a number of questions pertaining to head life and transport maintenance that are raised when discussing the amount of head cleaning agent. Regarding head life, it is frequently stated that the alumina added to a tape formulation is responsible for premature head deterioration. This is not necessarily the case. The amount of head cleaning agent added is extremely important. If too much alumina is added certainly the abrasivity of the tape is higher and head life deterioration becomes an issue. If too little alumina is added

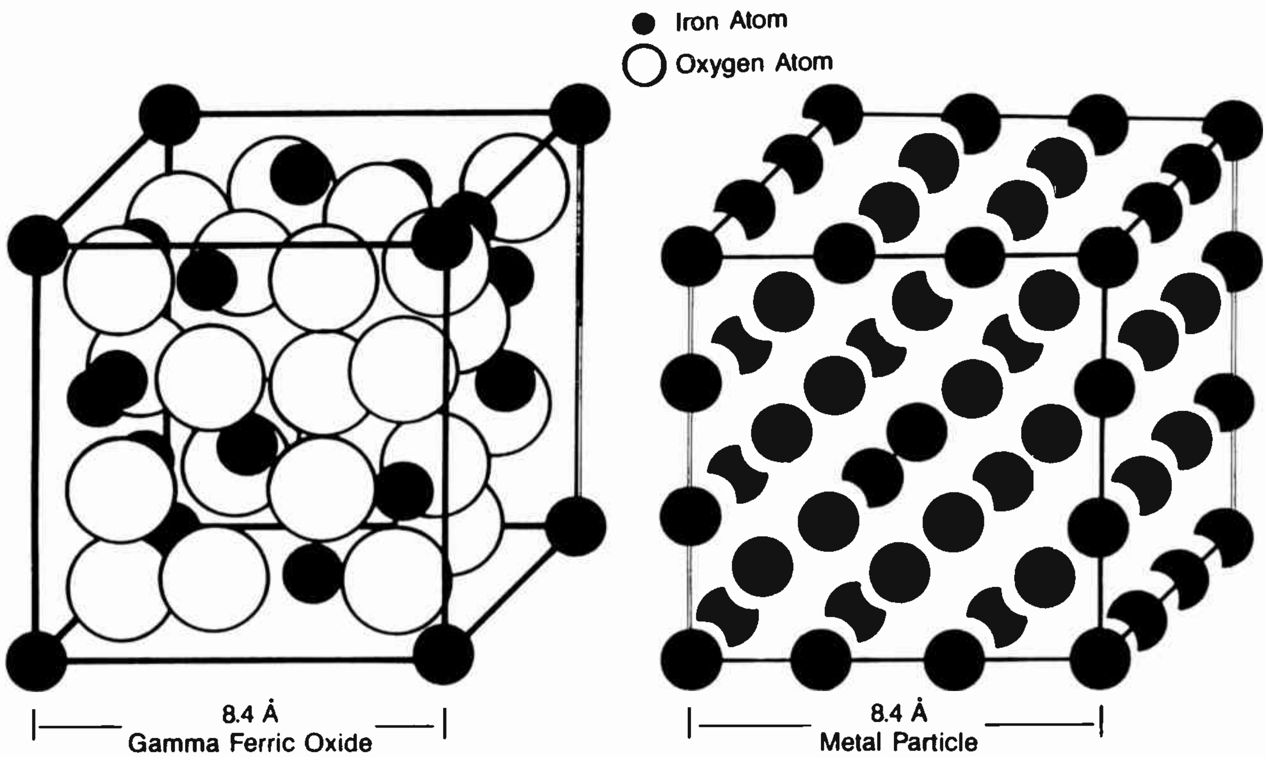


Figure 5.5-12. Crystal structures for metal iron and iron oxide.

the tape may not be able to keep the transport clean enough to ensure good head-to-tape contact and a poor quality recording results. There are many other factors that contribute to the actual abrasivity of a given tape. The type of particles, binder systems and other formulation ingredients all play a role in determining a particular tape's abrasivity.

The head cleaning agents added to a tape's formulation are not a substitute for good engineering practices in the form of routine cleaning and transport maintenance. Routine cleaning and alignment checking are part of any good preventative maintenance regime and will help to reduce the incidence of dropouts and machine induced physical damage.

Solvents

All of the mix ingredients are added together to form a slurry and part of the formulation will include some solvents. These are used to dissolve other mix chemicals such as binders during the milling process. The solvent system must maintain solubles in solution and not disturb the pigment solution. In addition, these solvents add to the formulation's rheological properties and effect the manner in which the finished mix flows. This directly impacts the manner in which the particular formulation performs on the coating line. While the solvents are needed during the mixing and coating of the formulation they are not required in the final state of the magnetic tape. The solvent laden air can be collected during the drying phase of manufacture and routed to a solvent recovery system. Here the solvents are recovered using a distillation process. This is an environmentally friendly process and can typically yield solvents of a purer quality than those originally purchased from suppliers and therefore reusable in another coating mix.

Dispersants

Because of their magnetic nature, particles tend to clump together in their raw form. This situation is not desirable as it would tend to result in some areas on the tape surface highly saturated with particles, whereas other areas may be entirely void of magnetic material. Dispersants are added to the formulation to aid in dispersing the particles evenly throughout the mix. The dispersion process is completed in mix preparation and can be achieved through a variety of methods. These methods are discussed in the subsequent section on mix preparations.

Fungicides

Fungus growth on the surface of tape may occur and is almost entirely confined to storage or operating environments with very high humidity. Fungicides are added to the mix formulation to help prevent fungus growth. Despite manufacturers' best efforts, poor storage environments are commonplace.

Antistatic Agents

One of the causes of dropouts (the momentary loss of head-to-tape intimacy) is static electricity buildup.

The buildup causes dust and dirt particles to attach themselves to the tape surface and cause a lack of head-to-tape intimacy. Antistatic agents are, therefore, added to tape formulation, particularly backcoat formulations. Carbon black is a common antistatic agent that is used for this purpose.

Mix Preparation

Once all of the formulation ingredients have been combined, it is necessary to mill or disperse the slurry so that two conditions are met. First, the magnetic particles, and other mix ingredients, must be thoroughly dispersed throughout the slurry. Second, all ingredients must be dissolved so that no particles larger than a given size remain. This is achieved in a milling process, which can be performed in a variety of ways. A typical milling cycle maybe from 72 to 96 hours and depends on the chemistry of the formulation. It is normal to filter the mix after milling. Filtering will remove unwanted materials such as foreign particles, (broken pieces of the milling media which were introduced during that process). At the end of the milling process, the slurry is ready for coating.

Ball or Pebble Mills

Ball mills are the oldest of the milling technologies. The ball mill is a horizontally oriented cylindrical rotating vessel into which the formulation is poured. The vessel is than rotated for a predetermined amount of time. The time is determined during the formulation development process. Inside the vessel is the milling media. In this case the media can either be steel balls or smooth pebbles. The media in the mill acts on the formulation ingredients causing them to be fully dispersed throughout the slurry.

Sand Mills

Sand mills are a more recent technology and operate on a flow system rather than the static batch system of the ball mills. Sand mills are actually banks of milling vessels vertically oriented. The sand mills themselves do not rotate, rather, the mix is pumped through each vessel from bottom to top. As the mix is pumped through the vessel it must pass through a series of rotating discs and milling media. The media in this case is typically zirconium silicate. The number of vessels through which the mix must pass is determined during the formulation development process.

High Solids Mixers

The most recent advances in milling technology are high solids mixers, the most common, double planetary mixers, can be likened to an industrial dough mixer in a bakery. High solids mixers can operate in either a batch mode, like balls mills, or a continuous flow mode, like sand mills. As the name implies, this type of technology requires a much more viscous mix than do ball or sand mills. Consequently, this technology provides an efficient and effective milling procedure. High energy kneaders are becoming very common for the high coercivity formulation of magnetic particle tape. This type of mixer is similar to a plastic injection molding screw drive. A number of passes is usually required through the kneader and the mix is very dense and paste-like compared to traditional mixes. This kneaded mix is then converted to a more traditional slurry consistency before coating.

Base Film

When magnetic tape was first coated it was constructed on kraft paper backing. This was later replaced with a more resilient cellulose acetate and then with today's range of polyester base films or backings. The current polyester base film includes Polyethylene Terephthalate (PET), Polyethelene Naphthalate (PEN) and Polyaramide type films.

Polyester type base film can be produced in a variety of widths, lengths, thicknesses, surface smoothness and tensile properties. Different magnetic recording systems call for magnetic tape products with different physical and tensile properties. The base film is electrically inert and provides no electrical or magnetic function. Base films are becoming thinner with newer tape formats. This presents a unique challenge to the tape manufacturer. It is essential that tensile properties are maintained despite the move toward thinner films. Base film derives a lot of its stiffness from its thickness. Obviously, the thinner the film, the less stiff the magnetic tape may be. Base film manufacturers are developing newer products to meet the changing requirements of the magnetic tape recording system. PEN and polyaramide base films are the next generation of films expected to someday take over from standard PET films. As a general rule of thumb, both PEN and polyaramide films provide greater tensile strength than PET film of the same thickness.

Magnetic Coating

Coating can be achieved by using a variety of coating technologies. As the requirements for higher magnetic tape has increased over the years, the coating technology to produce these newer tapes has been the result of this necessity. Older tape was thicker (base film and magnetic coat), coated at slower speeds and was relatively easier to coat. The coating technology to produce the thinner and smoother generation of metal particle tapes are more technologically complicated. However, each coating technology is valuable in that it does produce the desirable tape product to meet the requirements of its respective recording system.

Knife Coating

Knife coating is probably the oldest type of coating method employed in the manufacture of magnetic tape. As Figure 5.5-13 shows, the coating mix (wet) is fed onto the base film and then past the knife blade. The blade height (with respect to the base film) determines how much mix is left on the film. Knife coating is particularly good for relatively thick coatings such as those found on professional audio mastering tapes. This method is not the most accurate and is best for coating thickness above 400 μ in.

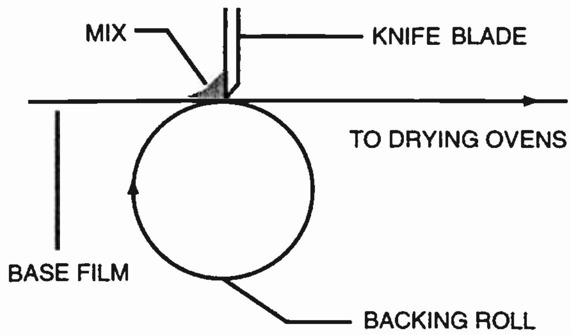


Figure 5.5-13. Application of knife coating.

Reverse Roll Coating

Reverse roll coating is more complex than knife coating and, consequently, is capable of coating much thinner coatings. As shown in Figure 5.5-14, in the reverse roll coating method, the mix is picked up from a mix pan by the pickup roll and transferred to the applicator roll. A metering roll than meters the amount of mix on the applicator roll. The amount of mix remaining on the applicator roll is determined by the distance between the two rolls. The mix is then transferred to the base film from the applicator roll. The applicator roll is turning in the reverse direction or against the motion of the base film, hence the name. Reverse roll coating was the mainstay of the tape industry for many years and is still in common use today. Reverse roll coating is better suited to thinner coatings (above 200 μ in.) such as those found on some older professional videotape formats process.

Gravure Coating

Gravure coating was adapted to the tape industry from the printing industry. It is the most accurate means of coating thin thickness where tight tolerance control is required. The gravure roll is pitted with cells, small grooves or cuts of given width or diameter and depth. The diameter and depth of the cell, as well as the viscosity of the mix, together determine how much mix is retained by the gravure roll and thereby deposited onto the base film. Gravure coating is an ideal method for very thin coatings such as consumer VHS and the backcoat coating found on most professional videotapes (see Figure 5.5-15).

Metal Evaporation Process

The process or coating technique of depositing a metal layer onto a flexible substrate through evaporation of a metal in a vacuum is referred to as metal evaporation (ME) coating. This process was developed to increase packing density, or meet the need to be able to record more and more data on the tape. Therefore, make the tape thinner and hotter (or more magnetic).

One means of increasing the volumetric packing density of magnetic tape is to make the magnetic layer thinner while preserving or increasing its magnetic capabilities. A means of achieving these seemingly contradictory objectives is to deposit a continuous pure metal film onto the base film, instead of using the typical dispersed particulate coating. In the introduction to this section the actual volume of magnetic oxide in typical coatings was reviewed. The most direct way to maximize the magnetic volume of the recording layer is to remove all non-magnetic components (oxygen) of the particle itself and by removing binding and resin matrix from the formulation. The ideal magnetic recording medium is achievable through this physical deposition of magnetic material on the base film through metal evaporation. The technique of depositing a metal on a flexible base film is based on the evaporation of a metal in a vacuum. As pressure is increased (vacuum increases) liquids evaporate more rapidly and at a lower temperature. A vacuum chamber used for metal evaporation onto a base film consists of the following components (see Figure 5.5-16):

- Transport system for unwinding and rewinding the base film

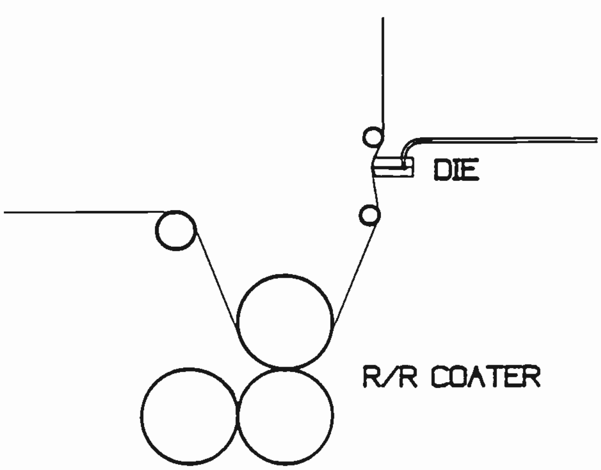


Figure 5.5-14. Application of reverse roll coating.

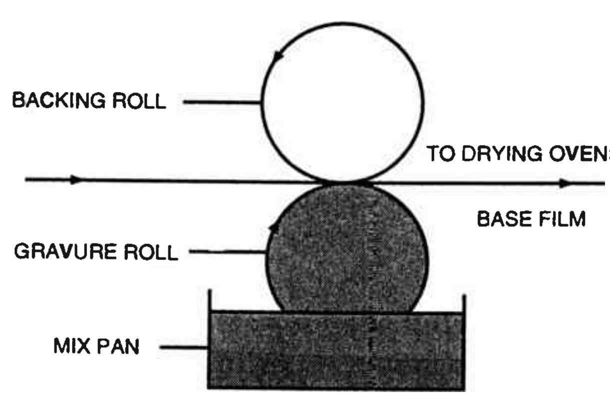


Figure 5.5-15. Application of gravure coating.

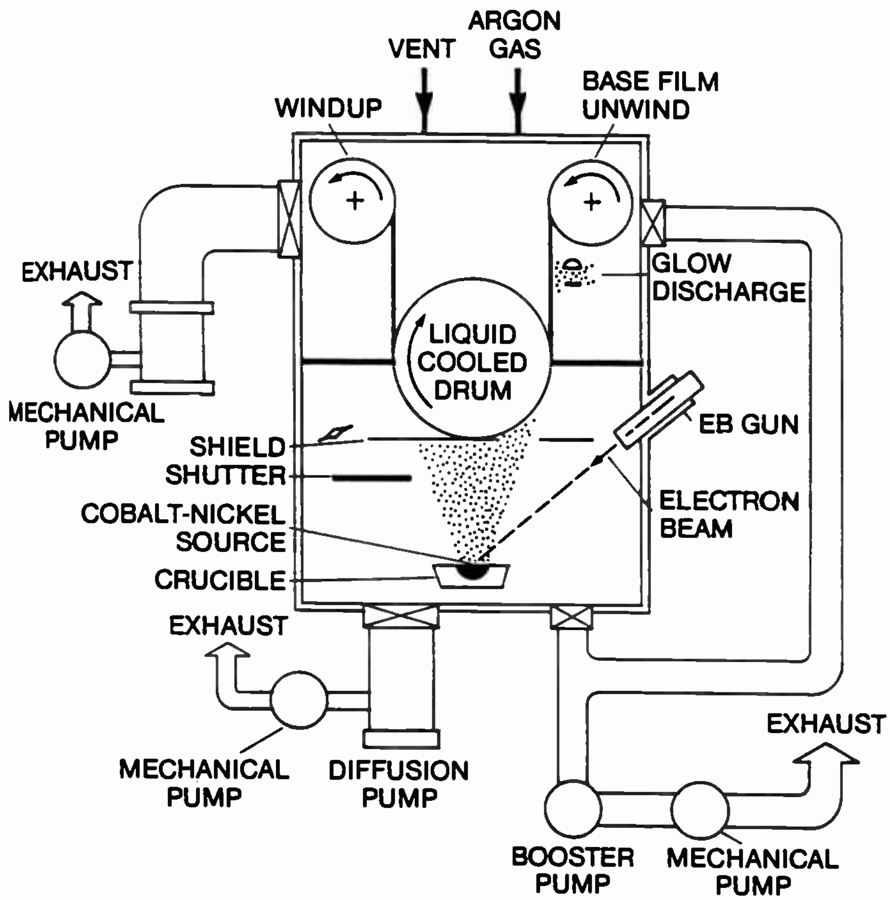


Figure 5.5-16. The vacuum deposition process.

- A high temperature crucible that can hold molten metal after it has been heated either by resistance or with an electron beam gun
- An internally chilled cooling drum around which the base film is wrapped to remove the heat from the condensing metal
- A vacuum pump to quickly remove the air and water vapor from the vacuum chamber

The magnetic coating thickness of a metal evaporated tape is typically 1 μ or 4 μ in. This is more than 20 times thinner than the metal particle layer of D-2 tape. This practically means that the entire thickness of the tape is basically that of the base film being used. The advanced magnetic properties of metal evaporated tape have allowed equipment manufacturers to significantly increase packing densities.

Slot Die Coating

As shown in Figure 5.5-17, slot die coating uses a coating head that has a slot cut into it. The geometry of this slot and its position relative to the base film determines the parameters of the coating surface. Slot die coating is one of the newest of coating technologies and is typically used for the thinnest and most critical of coatings. The slot die head usually has one slot but is capable of two or three slots which results in multiple and different layers being applied to the base film. These can be applied wet on wet or wet on dry.

Orientation

After leaving the coating area, the web, as it is now called, passes through the orientation process. This process physically moves the individual particles into a specific direction. The direction desired is always identical to the direction of head travel across the tape or as near to this direction as possible. In the case of the original videotape format quadruplex, developed by AMPEX in 1956, the head traveled in a lateral or transverse, direction. The particles in quad tape were therefore oriented in the lateral or transverse direction.

Drying, Solvent Evaporation, and Recovery

As noted previously, the mix formulation contains a number of different solvents. These solvents are undesirable in a finished tape and are therefore removed during the drying process by heating and causing evaporation. After leaving the orientation stage,

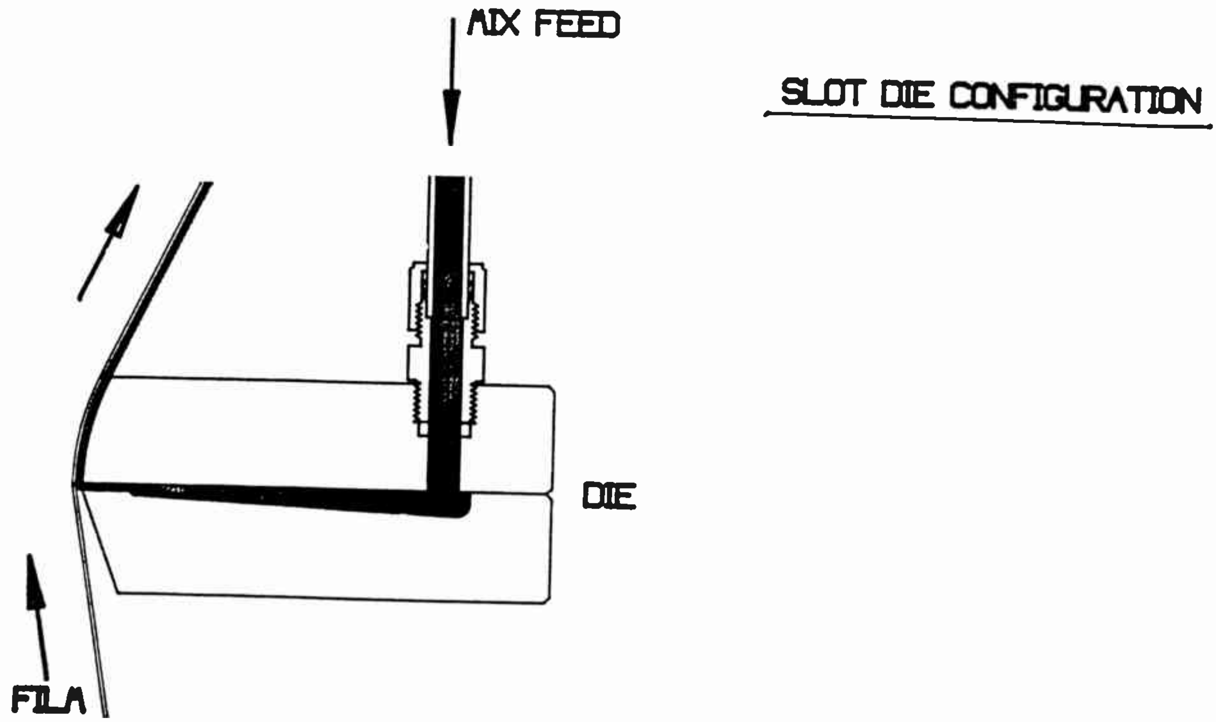


Figure 5.5-17. Slot die coating.

the coated web passes through an oven with a series of different temperature zones. The temperature of each zone is usually matched to the evaporation temperature of one of the solvents being used. Within the ovens, the web typically travels over a series of air plenums, which suspend the tape on a cushion of hot air. This prevents the tape from touching any of the surfaces inside the oven, which would damage the coated surface.

The oven air is gathered in the process and is sent to the solvent recovery unit. This is a relatively simple process whereby the air is passed through a vessel containing activated charcoal. The charcoal absorbs the evaporated solvent and the air is passed into the atmosphere. The charcoal is then flushed with steam and the solvents recovered as part of a distillation process.

Calendering

After passing through the ovens, the coated web then passes through a series of calenders. The coating layer is usually done at a thickness a few times more than actually desired in the finished product. This is to compensate for shrinkage during the drying and calendering processes. The calendering process is a smoothing or surface finishing process that is accomplished by heat and pressure. Calendering may either be performed inline, that is, in a continuous process, or offline, which involves a separate set of calender stacks to those found on the coating line itself.

A calender stack consists of two rolls located side by side. The meeting point of the two rolls is called a nip. One of the rolls provides pressure and the other heat. The calendering process can be roughly likened to ironing of clothing. During the process, which usually consists of a number of nips, the tape surface is smoothed and compressed. The smoother tape surface tends to give higher electrical outputs than rougher ones and better signal to noise ratio.

Take Up

If the calendering process is performed inline, the last task in the coating line is to wind the coated and calendered web onto a web core ready for the slitting process. This station acts in the very same manner as the take up reel on a reel-to-reel recorder.

Slitting

A 26 in. or 3 meter wide jumbo roll of tape in roll form is useless to anybody. It must be slit and packaged into a usable form. Slitting takes the jumbo roll and cuts it into strands of the appropriate width for the given format. Each strand is then wound onto a hub. This is commonly referred to as a pancake. All slitters operate in a similar manner. The web passes through a series of blades, known as arbors. There are two major parts to the arbor. The first is the male cutting blade, which is typically positioned above the web surface. The second is the female or guides, which are typically positioned below the surface.

The male and female portions of the arbor operate together to form the slitting process. Edge weave and edge quality are all critical factors that must be tightly controlled during this process. Today's thinner magnetic tape formats are becoming more demanding on the integrity of the magnetic surface. As well as the tape surface itself, edge quality becomes a major issue. Defects in edge quality are greatly magnified in smaller formats, which depend on every mechanical component of the recording process being within a very small tolerance range in guiding and tracking.

Curing

A process that is required for most crosslinked binder systems of metal particle formulation tape is heat curing. The heat curing can occur at different phases of the manufacturing process, before the back coat is applied or after the entire product has been coated. Heat curing can be done on full webs or on slit pancakes. It is usually done at a temperature cycle and for a period of time determined to be optimum for that specific formulation. The heat curing phase can take a few hours or for a few days.

Post Conditioning

Also referred to as burnishing, the post conditioning phase is used to condition the edges and to clean the tape after it has been slit to pancakes for the specific format. The burnisher is typically similar to a tape transport and can consist of different type of cleaning and wiping materials. The number of burnishing sections, the type of blade or device used in each and the tape tension can be adjusted for the particular formulation.

Assembly

The final task in the magnetic tape manufacturing process is to load and assemble the cassettes or to add flanges to hubs. In cassette format products, manufacturers have a choice between spool loading and in-cassette loading. Spool loading is a process where the spools are wound with the correct amount of tape prior to being placed into the unfinished cassettes. The cassette is then assembled around the spools. In-cassette loading assembles the cassette first, and then winds the footage into the cassette by cutting and splicing the leader material from outside of the cassette. Large scale assembly of cassettes and their component parts is largely achieved using relatively simple robotics and industrial controls. Typically, assembly lines have several stations where a certain task is performed: placing a guide, placing a break lever, inserting and driving screws into final position, etc. The use of robotics helps to improve consistency of product and lower the overall cost of manufacture. The assembled cassette or flanged open reel, is then placed in the appropriate unit container, labeled, and placed in the shipping carton and ready for sale.

BIBLIOGRAPHY

Lowman C., Magnetic Recording, McGraw Hill Inc. 1972.

Hammar, P. and Osoke D., The Birth of the German Magnetophone Tape Recorder 1928–1945, dB March 1982.

Jorgensen, F., The Complete Handbook of Magnetic Recording, 3rd edition., McGraw-Hill Inc., 1988.

Hadad, A., Metal Evaporated Tape Technology, Post, May 1988.

Kempster, L., Media Mania 5th ed, Avedon Associate Inc., 1997.

Mee, C. and Daniel, E., Magnetic Recording Handbook, Technology & Applications. McGraw-Hill, 1987.

Pohlmann, K., "MD," Mix, August 1991.

Pohlmann, K., "Mini Disc Technology," Mix, December 1992.

Stafford, D., "Continuing Developments in Magnetic Tape Technology," Broadcast Engineering, October 1994.

Watkinson, J., The D-2 Digital Video Recorder, Focal Press, 1990.

Williams, E.W., The CD-ROM and Optical Disc Recording Systems, Oxford University Press, 1994.

KEY TERMS

Ablative. A method of optical recording where holes are made in the optical media. Also referred to as pit forming. A laser actually burns a hole into the film layer. Data is read by reflectivity contrast (WORM)

Birefringence. The property of a material which causes incident light waves of different polarization to be refracted differently by the material.

Coercivity. A measurement of the magnetic characteristics of a given particle. Coercivity refers to the amount of magnetic energy required to bring a saturated magnetic particle to a zero or demagnetized, state.

Curie point (or Curie temperature). The temperature at or above which the coercivity of magnetic material decreases substantially, permitting a change in orientation by a weaker magnetic field.

Embossed. Data written marks are molded into the substrate. Data is read by reflectivity contrast (ROM).

Kerr effect. The phase change exhibited by certain substances when exposed to a magnetic field, due to rotation of the plane of polarization of light reflected from them.

Laser. A device that uses the natural oscillations of atoms to amplify or generate electromagnetic waves into an intense beam of coherent optical radiation. (Acronym for light amplification of stimulated emission of radiation.)

Phase change (PC). An amorphous coating in which individual spots on the disc are changed by laser irradiation from a crystalline state, which reflects light, to a non-crystalline state, which absorbs light. Data is read by reflectivity (WORM or Rewritable).

Reflectance (reflectivity). The ratio of the incident light to the reflected light measured at normal incidence on the disc at a specified wavelength, through the protective layer, if any.

5.6 VIDEO COMPRESSION SYSTEMS

PETER SYMES
TEKTRONIX, GRASS VALLEY PRODUCTS, GRASS VALLEY, CA

This article is based on material extracted and adapted from the book Video Compression *by Peter Symes and published by McGraw-Hill (ISBN 0-07-063344-4), by permission of the publisher.*

INTRODUCTION

Compression is the science of reducing the amount of data used to convey information. This article examines a wide range of techniques—some simple, some very complex—and how these techniques may be applied to video. Compression relies on the fact that information, by its very nature, is not random but exhibits order and patterning. If this order and patterning is extracted, the information can often be transmitted using less data than would be needed for the original. The original, or a close approximation to it, can then be reconstructed at the receiving point.

It may be tempting for the television engineer to conclude that compression is a subject for experts, but good engineers understand in principle what is happening in each of the major components of a system, even if they could not design that component. Compression will exist at some point in virtually all television systems and its performance is affected, sometimes dramatically, by characteristics of the video signal including any prior compression history. An understanding of compression principles will help video engineers to understand the processes that are being applied to their images. Only in this way can they protect quality and diagnose problems.

There are several families of compression techniques, fundamentally different in their approach to the problem—so different in fact that often they can be used sequentially to good advantage. Sophisticated compression systems use one or more techniques from each family to achieve the greatest possible reduction in data. The JPEG and MPEG standards are the most important today; many systems already installed use JPEG or its derivatives, new systems will mostly be based upon MPEG-2.

One important categorization of compression systems is the degree of *symmetry* or *asymmetry*. For applications such as video conferencing where there are similar numbers of transmitters and receivers, there is no particular advantage if the decoder is less expensive than the encoder. For other applications such as broadcast a very small number of encoders generate signals that will be sent to a large number of decoders. Such systems can be asymmetric; the decoder must be as inexpensive as practicable, but the encoder can be very complex if this improves efficiency.

INFORMATION AND DATA

Information exists in many forms—written text, the spoken word, music, still pictures, and moving pictures are just a few examples. Whatever the type of information, we can represent it by electrical signals or data to transmit, or store it.

The more complex the information, the more data is needed to represent it. Plain text can be represented by eight bits per character, or about 20 kb for a page. CD quality music requires nearly 1500 kb for each second, and full motion 525-line component video needs over 200 Mb for each second.

In order to transmit a given amount of data, a certain data bandwidth (measured in bits per second) is used for the time necessary to transmit all the bits. In the real world, there is always some restriction on the available bandwidth. For example, if the transmission is sent over regular telephone lines using modems it is difficult to achieve more than 33.6 kb/s (kilobits per second). This is fast compared to the 300 b/s modems of a few years ago, but even at 33.6 kb/s it would take two hours to transmit a single second of high quality video.

Sometimes it is a practical restriction like the use of a modem that sets the requirements for compression. It is possible to send high quality uncompressed audio data over a modem, but each second of audio would take about 50 seconds to transmit. The data would have to be received gradually and stored away; then the resulting file could be played at the correct rate to hear the sound. However, it may be important to transmit sound so that we can listen to it as it arrives in "real time." Quality reduction would help, but even telephone speech quality needs a data rate substantially higher than a high-speed modem can provide. To send real-time audio over a modem link, the data has to be compressed.

Another example of this type of requirement is the digital television system developed for North America. When the study was started in 1987, few thought it would be possible to transmit high definition television (characterized by the 1125-line system) over a single 6 MHz channel as used for today's 525-line television. When digital techniques were considered, the problem looked even worse, as the high definition signal represented over 1 Gb/s of data. In the end, two teams of engineers combined to produce the answer; the transmission engineers designed a system to reliably deliver nearly 20 Mb/s over a 6 MHz channel, and the compression engineers achieved data reduction of about sixty to one, while maintaining excellent picture quality.

Sometimes the motive for compression is purely financial. Bandwidth costs money—compression reduces the cost of transmission. A satellite link with a data bandwidth of 40 Mb/s will earn more money if it can carry ten television channels instead of just one.

Transmission and storage of information may seem to be very different problems, but when considering the quantity of data, the economics are the same. Just as there is a cost for each bit transmitted, there is a cost for each bit stored in memory, disk or tape. If less data is used, both transmission and storage will be cheaper. There are some differences in how compressed data is best structured, but otherwise everything said about compression applies equally to transmission or storage.

Signal Conditioning

It is extremely important to ensure that the signal or data stream to be compressed represents the information that must be transmitted, and *only* that information. If there is additional information it will require bits to transmit, and result in fewer bits being available for the necessary information. Surplus information is irrelevant because the intended recipients can make no use of it.

Surplus information can take many forms. It can be information in the original signal or data stream that exceeds the capabilities of the receiving device. For example, there is no point in transmitting more resolution than the receiving device can use.

Another form of surplus information is artifacts—features or elements of the input that are not truly part of the information. Noise is the most obvious example. Noise is by nature random or nearly so and this makes it essentially incompressible.

Many other types of artifacts exist, ranging from filter ringing to film scratches. Some may seem trivial, but in the field of compression they can be very important. Compression relies on order and self-consistency in a signal, and artifacts damage or destroy this order.

This statement is just one aspect of a more general rule. Compression systems are designed for particular tasks and make use of assumptions about the nature of the data being compressed.

Video is generally an environment where images are sampled according to the limitations of the *Nyquist Theorem*. The theorem can be expressed as:

> *To ensure that a signal can be recovered from a series of samples, the sampling frequency must be more than double the highest frequency in the signal.*

The frequency transforms widely used in video compression, particularly the *Discrete Cosine Transform,* depend for their action on the fact that the information is band limited, and that sampling was performed according to Nyquist. If these rules are broken the transform does not perform as expected, and the compression system will fail.

Correctly used, signal conditioning can provide remarkable increases in efficiency at minimal cost. Equipment available today incorporates many different types of filters targeted at different types of artifact. The benefit of appropriate conditioning is twofold. Because the artifacts are unwanted, there is a clear advantage in avoiding the use of bits to transmit them. Moreover, because the artifacts do not belong they generally break the rules or assumptions of the compression system. For this reason artifacts do not compress well and use a disproportionately high number of bits to transmit.

Lossless Compression

Some compression techniques are truly lossless. In other words, when data is compressed, the process can be reversed (decompression) and the original data can be retrieved. Lossless compression works by removing *redundant* information—information that, if removed, can be recreated from the remaining data.

In one sense this is the ideal form of compression—there is no cost to using it other than the cost of the compression and decompression processes. Unfortunately, lossless compression has two significant disadvantages. Typically, lossless compression offers only relatively small compression ratios, so used alone it often does not meet the economic needs. Also, the compression ratio is very dependent on the input data. Used alone, lossless compression cannot guarantee a constant output data rate such as might be required for a transmission channel.

For some applications lossless is the only type of compression that can be used. If a binary file such as a computer program is to be transmitted, receiving an approximation to the original is of no use. The program will only run if replicated exactly. Fortunately, many computer files have a high degree of order and patterning, so lossless compression techniques do yield very useful results. Compression programs like the well-known PKZIP use a combination of lossless techniques, and are particularly effective on graphic image files.

One advantage of lossless compression is it can be applied to any data stream. Most video compression schemes use lossy techniques to achieve large degrees of compression. Lossless techniques are then applied to the resulting data stream to reduce the data rate even further.

The simplest form of lossless compression is *run-length encoding*. Many data streams contain long runs of a single value. For example, a graphics file will typically contain values for each pixel, ordered line by line. A red object will likely be represented by several long runs of the value red. With appropriate coding it can use much less data to say *Red 23 times* instead of "R R R R R R R R R R R R R R R R R R R R R R R." This is run length encoding—the code represents the length of the run of a certain value. As this chapter will describe, various tricks are used in video compression to make the data suitable for run length encoding.

A more sophisticated type of lossless compression is *entropy encoding,* also know as *variable length encoding*. This is usually the last step in compression. Where run length encoding relies on adjacent values being the same, entropy encoding looks at the overall frequency of specific values, wherever they are located. The implementation of entropy encoding is quite complex, but the principle is easy to understand.

If the data is represented by bytes (groups of 8 bits), each byte can have any *value* between 0 and 255. When the data is transmitted a *symbol* is used to represent each possible value. A sequence of *symbols* is transmitted then until all the *values* have been sent.

The simplest way to code the data is to send the bytes themselves—each symbol is the same as the value it represents. This is so obvious that there is a tendency not to think of anything else, and do not distinguish between the value and the symbol; but it does not have to be like this. As long as both the transmitting and receiving points know and apply the same set of rules, any set of 256 unique symbols can be used to represent the 256 possible values. This is not very helpful without an added factor—provided there is a set of rules that lets the recipient know when one symbol ends and another begins, *not all the symbols need to be the same length.*

This is the essence of entropy encoding. It compresses the data by using short symbols to represent values that occur frequently, and longer symbols to represent the values that occur less frequently. Unless the input data is very close to random in its distribution of values, this technique means that total data used for the symbols will be less than the total of the original values.

Optimization of entropy encoding is complex, but the basic idea is neither difficult nor new. Consider two ways of coding the alphabet. Anyone who has spent much time with computers has come across the ASCII encoding system. In ASCII, letters are represented by a particular set of eight bits, or one byte. It does not matter what the *value* (letter), the *symbol* is eight bits long. A much earlier method of encoding the alphabet was invented by Samuel Morse in 1838. The *Morse Code* is a simple but useful example of an entropy coding system.

The Morse code uses combinations of dots and dashes as symbols to represent letters. A dash is three times as long as a dot; each dot or dash within a symbol is separated by one dot length; each symbol is separated by one dash length. Morse designed his code to transmit English language text. To maximize the speed of transmission he looked at the frequency with which the various letters of the alphabet appear in typical text. The letter "e" is the most common letter in English, so Morse assigned the shortest possible symbol—one dot (actually four dots in length because of the inter-symbol space). The next most common letter "t" received the next shortest symbol, one dash. At the other extreme, an infrequent letter like "y" receives three dashes and a dot, a symbol four times as long as that used for "e," as shown in Figure 5.6-1.

The Morse Code also illustrates one of the problems of entropy encoding. For transmission to be successful, obviously both the transmitter and the receiver must use the same mapping of symbols to values. The relationship between values and symbols is known as the *code table.* But the most efficient mapping can only be determined when the frequency of values is known. The code table for Morse is derived from the frequency of letters in English text. This code would not be efficient to send Polish language text, where the letters "c" "y" and "z" are used much more frequently than in English because those letters have long symbols. This is probably not a major concern when transmitting a few lines of Polish in the middle of many pages of English, but it would be advisable to change the code table before transmitting a large volume of Polish text. In order to do this rules must be established at the transmitter and receiver to allow for different code tables. However if the code tables are changed, the new table must be transmitted and that means transmitting a considerable amount of additional data!

There is no perfect solution to this, and the best compromise depends on the application. Some coding schemes use only one code table. This may not always be the most efficient table, but the coding system does not carry any of the overhead necessary to change tables. Other systems have a small number of fixed code tables, known to both transmitter and receiver. The encoder can examine the data to be transmitted, and select the most efficient code table. The only overhead is the provision of a message that tells the receiver to use, say, "table number three." Still other schemes allow for the transmission of custom code tables whenever the characteristics of the data change substantially.

In 1952 Huffman demonstrated a method for deriving an efficient set of (variable length) symbols for a data source with known statistics. This method is widely used today, and entropy coding is frequently referred to as Huffman coding.

Figure 5.6-1. Morse Code provides a simple example of variable length encoding.

Lossy Compression

As mentioned in the previous section, lossless compression would be an ideal answer, except that it rarely provides large degrees of compression, and it cannot be used alone to guarantee a fixed bit rate. Lossless compression is an important part of the tool kit, but on its own it does not provide a solution to many practical problems. To put digital audio over a modem link, or high definition television through a 6 MHz channel, some loss from the compression system must be tolerated—what comes out will not be exactly the same as what went in. This is the field of lossy compression. Ideally lossy compression, like signal conditioning, removes *irrelevant* information. Some information is truly irrelevant in that the intended recipient cannot perceive that it is missing. There is also information that is close to irrelevant, where the quality loss is small compared to the data savings.

The objective of lossy compression is simple—maximum benefit (compression ratio, or bit rate reduction) at minimum cost (the loss in quality). However, the realization of this objective is not simple. A very large number of parameters have to be chosen for any given implementation, and many of these parameters must be varied according to the dynamic characteristics of the data. Even with a simple measure of the resulting quality (or lack of it), optimizing such a large number of variables is a very complex task.

Unfortunately there is as yet no simple measure that can be applied to the quality of a compression system, particularly for video. The only measure that really matters is the subjective effect as perceived by a viewer, and this is a very complex function. If compression loss were confined to one characteristic of a picture, it would be relatively simple to derive a way of measuring this characteristic objectively. Subjective tests could be performed on a representative sample of viewers to arrive at a calibrated relationship between this characteristic and subjective quality. Unfortunately, compression loss results in not one but many changes to the characteristics of the picture; each of these has a complex, non-linear relationship to subjective quality and they interact with each other.

Compression loss in images involves two main components: things that should be in the picture, but are lost; and things that are added to the picture (artifacts) that should not be there. Expected areas of a compression-related loss include spatial resolution (luminance and color, probably to different degrees) and shadow and highlight detail. Where extreme compression is needed for applications like video conferencing, or the Internet, temporal resolution would also be sacrificed. Artifacts that might be added include blocking, *mosquito noise* on edges, quantization noise, stepping of gray scales, patterning, ringing, etc. As a further complication, in the presence of a loss, the addition of small amounts of artifact such as noise or ringing may actually improve the subjective quality of the picture.

Even within standardized compression schemes such as MPEG, the tuning of a system to yield the maximum quality-per-bit is a black art involving models of the human psychovisual system that are closely guarded commercial secrets. Evaluation of the effect of parameter changes is very complex for the reasons discussed above.

Quantization

For lossy compression, information in some form has to be discarded, and quantization is the tool used most frequently. Quantization determines the precision with which values are represented. As previously mentioned, information needs to be represented with sufficient precision, but it is important not to transmit unnecessary information. Properly applied, quantization can ensure that the information transmitted or stored has just enough precision for the intended application. Quantization is also used to reduce information beyond this point, knowing that impairments will be introduced. Part of the job of a compression system is to ensure that the data is arranged so that the information discarded by quantization has the minimum possible subjective effect on the delivered image.

The simplest form of quantizer is the *scalar quantizer*, and the most common implementation is the *uniform scalar quantizer*. It operates on a one-dimensional variable such as intensity and corresponds to a staircase function with equal spacing of the steps. A continuum of input values is divided into a number of ranges of equal size. Figure 5.6-2 shows a range of input values from 0 to 255, divided by a uniform scalar quantizer into eight equal regions. There are 256 possible input values, but only eight possible output symbols.

As the individual input values are processed by the quantizer, they are compared with *decision values* representing the steps of the staircase. Any input value between the first two decision levels will be assigned the first output symbol, and so on.

The other parameter that has to be considered is the *reconstruction value*. When the quantizer shown in Figure 5.6-2 is applied there are only eight possible output values, each of these can be represented by a three bit code, or a value between zero and seven, as

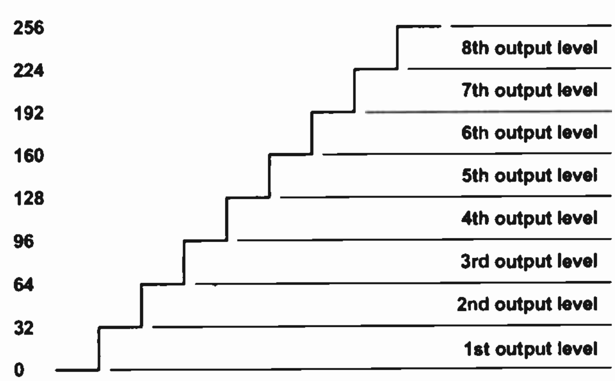


Figure 5.6-2. Uniform scalar quantizer.

Table 5.6-1
Choice of reconstruction value, and its effect on mean square error.

Input Range	Output Code	Reconstruction Value	MSE	Reconstruction Value	MSE
0–31	000	00000000 (0)		00010000 (16)	
32–63	001	00100000 (32)		00110000 (48)	
64–95	010	01000000 (64)		01010000 (80)	
96–127	011	01100000 (96)	325.5	01110000 (112)	85.5
128–159	100	10000000 (128)		10010000 (144)	
160–191	101	10100000 (160)		10110000 (176)	
192–223	110	11000000 (192)		11010000 (208)	
224–255	111	11100000 (224)		11110000 (240)	

shown in Table 5.6-1. When using the quantized data, it is no use applying intensity values zero to seven—these are all almost black. To get back to an approximation of the original range of input values, the quantized levels need to be a single value in the range 0 to 255. Adding five zeros to each three-bit symbol, would make all of the output values less than the input values but all of the errors would be in the same direction.

It is reasonably intuitive that if the input values are evenly distributed, a reconstruction value at the center of each step is likely to give smaller overall errors. This is illustrated in Table 5.6-1 where the overall error is expressed as a *mean square error* (MSE).

The most obvious application of quantization is to directly reduce the precision of the intensity values of the image. The necessary precision depends upon the application. Too few bits will produce a *paint-by-numbers* effect because the boundaries between adjacent values will be clearly visible. Assuming all the quantization steps are equally visible, the eye can detect intensity differences of about 1%. To avoid visible boundaries the quantization step should be smaller than 1%, preferably by a factor of about two to one. To achieve a contrast ratio in a displayed image of 100:1 (about the best that can be achieved in a movie theater) the 256 values offered by 8-bit quantization should be adequate. For applications with lower contrast ratio, such as the printed page, a coarser quantization such as 6 bits may be sufficient.

Data Manipulation

Earlier the two most common methods of lossless compression, run-length encoding and variable length encoding were described. Reducing the amount of information by the use of quantization has also been covered. These techniques may all be used directly on image data, but far greater efficiency may be obtained if the form of the data is manipulated to take maximum advantage of the techniques. Two manipulation techniques commonly used in compression systems are *prediction* and *transformation*.

Predictive coding relies on the fact that continuous tone images have a high degree of correlation between the values of nearby pixels. In other words, gradual shading is more common than abrupt changes in intensity. Various prediction schemes may be used; the simplest is just to use the preceding pixel according to the scanning system used. The technique predicts or guesses each pixel value according to a known rule, then transmits just the error between predicted and actual values. At the receiving end, the decoder uses the same rule to make the same prediction, and uses the transmitted error value to obtain the correct value.

The range of possible errors is actually twice as large as range of intensity values, so initially this may not appear to be a good technique to use in compression. The secret lies in the statistical distribution of values. Intensity values in a typical image are scattered fairly randomly across all or most permissible values. However, if a good predictor is chosen most of the error values will be small—low values will be much more common than high values. This distribution is well suited to variable length encoding. In practice a typical continuous-tone image may be transmitted losslessly with about half the number of bits per pixel. If the error signal is quantized, then variable length encoded, an image may be transmitted to any desired quality level with fewer bits than would be needed for the same quality using direct quantization.

Predictive encoding is used in the lossless mode of JPEG, and is used in most compression schemes to compress sequences of correlated values such as dc coefficients and motion vectors.

Another powerful means of data manipulation is the *transform,* and the transform most commonly used in today's compression systems is the *discrete cosine transform* (DCT). DCT takes a block of (typically) 8 × 8 pixel intensity values and transforms these into an 8 × 8 block of DCT coefficients, representing a range of horizontal, vertical, and diagonal spatial frequencies within the original block. The DCT transform can represent any 8 × 8 block of pixel values by an appropriate mix of the 64 basis patterns shown in Figure 5.6-3. The appropriate mix is the 8 × 8 array of DCT coefficients. The process is reversible given sufficient arithmetic precision, but this requires 11 bit DCT coefficients to represent 8 bit intensity values. Again, the first step has increased the amount of data.

The benefit of DCT comes from a remarkable correspondence with the sensitivity of the human psychovisual system. When a continuous tone image is transformed by DCT, most of the energy is represented by a few high value coefficients. Many of the coefficients will be close to zero, and if these coefficients are forced to zero the effect on the reconstructed image to a human observer is negligible. Furthermore, it is found that the higher frequency coefficients may be very coarsely quantized with only small effects on the quality of the reconstructed image. The following description of JPEG will show how quantized DCT coefficients may be manipulated to make them well suited for run length encoding and variable length encoding.

JPEG

The Photographic Experts Group of the International Standards Organization began work on a data compres-

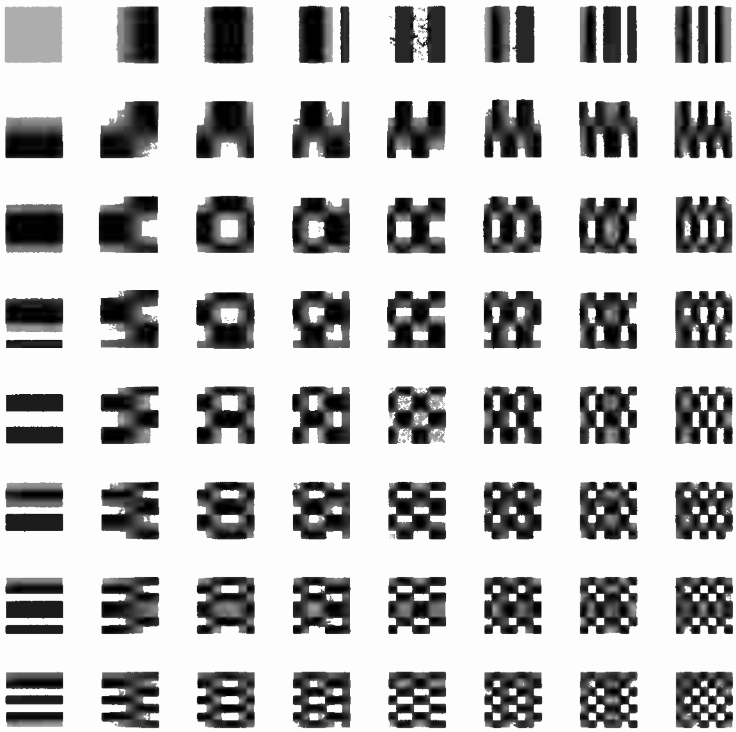


Figure 5.6-3. The DCT basis functions.

sion standard for color images as early as 1982. This effort was joined in 1986 by a study group from CCITT (now ITU-T) and became the Joint Photographic Experts Group (JPEG). In 1987 ISO and the International Electrotechnical Commission (IEC) formed a Joint Technical Committee for information technology, and JPEG continued to operate under this committee. Eventually the JPEG Standard was published both as an ISO International Standard and as a CCITT Recommendation.

At an early stage it was decided that the core JPEG compression would address monochrome images, and that compression would be applied separately to the various components of a color image. Given an 8 bit per pixel monochrome image, the original targets were to provide *recognizable* pictures at 0.25 bits per pixel, *excellent* quality at 1.0 bits per pixel, and images *indistinguishable* from the original at 4.0 bits per pixel. As the work proceeded the committee aggressively pushed performance requirements and the targets were revised several times. Final testing was performed against targets of 0.083 (recognizable), 0.75 (excellent), and 2.25 (indistinguishable) bits per pixel, and an addition level of 0.25 bits per pixel was added for *useful* quality images.

Many compression schemes were tested, but eventually DCT was chosen as the core technology for JPEG. DCT provided the best pictures at low bit rates, and is relatively easy to implement with fast algorithms that can be built into hardware.

This chapter covers JPEG only superficially, as it relates to the various tools already discussed. Because JPEG uses tools such as DCT, it is intended for applications that obey the assumptions made by these tools. It is intended for the compression of continuous tone images, *photographic* or *real-world* images as they have been characterized in this chapter; images that do not conform to these constraints can suffer badly. Specifically, JPEG is not designed for binary (black/white) images, nor is it well suited to discontinuous images such as limited color palette images (GIF images).

JPEG is a very extensive standard offering a large number of options for both lossless and lossy compression. This chapter addresses only the best known JPEG implementation—*Baseline JPEG*. This method uses DCT compression of 8 × 8 blocks of pixels, followed by quantization of the DCT coefficients, and entropy encoding of the result.

Baseline JPEG requires that an image be coded as 8 bit values. Compression is specified in terms of one value per pixel; color images are encoded and handled as (usually) three sets of data input. The JPEG Standard is color blind, in that there is no specification of how the color encoding is performed. The data sets may be *R, G, & B*, or some form of luminance plus color difference encoding, or any other coding appropriate to the application.

The most common coding used with JPEG is luminance plus C_B and C_R. C_B and C_R are derived from *(B-Y)* and *(R-Y)* respectively by scaling and shifting so that the full gamut for permissible *RGB* values can also be represented by 8 bit values. Level 128 represents the zero color difference. Usually, the C_B and C_R signals are filtered to half the spatial bandwidth of the luminance signal, so that each may be represented by only one quarter of the samples of the luminance (half the sampling rate horizontally and half vertically).

The first step in JPEG encoding is to divide the data into its different components, for example, Y, C_B, and C_R. Although the compressed data derived from the components may be interleaved for transmission (reducing the need for buffering at both ends), JPEG processes the various components quite independently. If the color components are bandwidth reduced, the effect is merely that the coder operates on a smaller picture. For example, a Y, C_B, C_R 4:2:2 picture might be processed as one 720 × 480 image (luminance) and two 360 × 480 images (CB and C_R). A simplified block diagram of both encoder and decoder for a single image is shown in Figure 5.6-4.

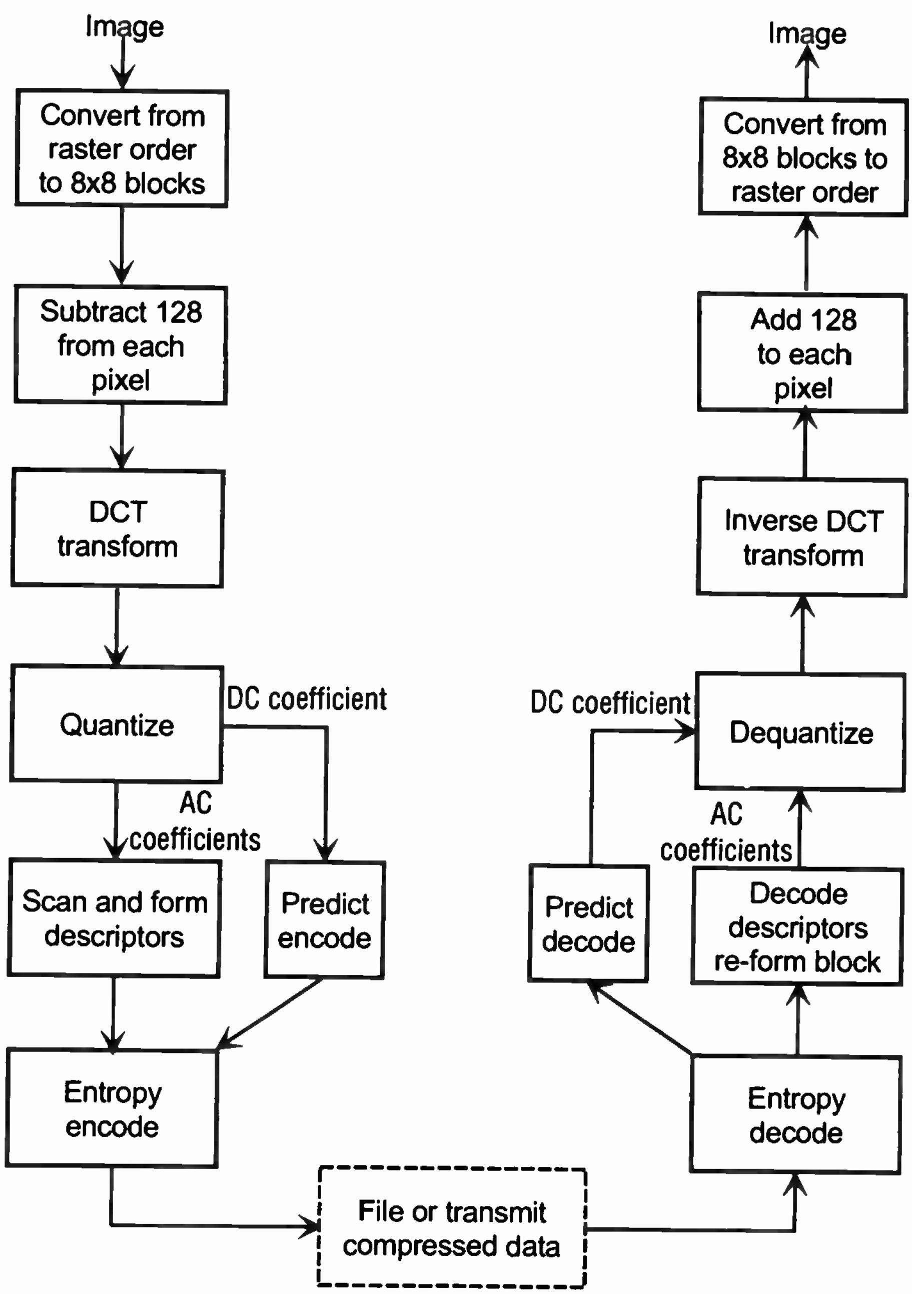


Figure 5.6-4. JPEG encode and decode.

Each component (and from now on we shall consider only one) is divided into 8 × 8 blocks, the value 128 is subtracted from each pixel, and each block is DCT transformed. The Forward DCT equation as used by JPEG is:

$$F(u, v) = \frac{1}{4} C_u C_v \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos \left(\frac{(2x + 1)\mu\pi}{16}\right) \cos \left(\frac{(2y + 1)\nu\pi}{16}\right)$$

where:

$$C_u = C_v = \frac{1}{\sqrt{2}} \textit{ for } u, v = 0,$$

$$C_u = C_v = 1 \textit{ otherwise}$$

This formula is not as daunting as it may appear, but it does involve a large number of calculations. $F(u, v)$ is the 8 × 8 array of DCT coefficients; $f(x, y)$ is the 8 × 8 array of pixel values. A single DCT coefficient (at location u, v) is calculated by summing the contribution from each of the 64 pixels (at locations x, y). This process is repeated 64 times to form the complete array of DCT coefficients. With pixel values, $f(x, y)$ in the normal range of 0–255, this equation yields ac coefficients in the range −1023 to +1023, and these may each be represented by a an 11 bit signed integer. The dc coefficient, however, would be in the range 0–2040, and this would require higher precision in the hardware or software performing the DCT transform. To avoid this discrepancy, the value 128 is subtracted from every pixel value prior to the DCT transform. This subtraction has no effect on the ac coefficients generated, but shifts the dc coefficient into the same range, so that all coefficients may be represented by 11 bit signed integers.

The next step is *quantization*. JPEG quantization is a sophisticated process that can effectively use a different quantizer for each coefficient in the 8 × 8 array. Higher spatial frequencies are quantized more coarsely than low frequencies. This process is one of the most significant to the power of JPEG.

The quantized dc coefficient (top left) is separated from the ac coefficients, and the sequence of dc coefficients (one from each block of pixels) is predictively encoded, and then entropy encoded (see Figure 5.6-4).

The quantized ac coefficients from a block are scanned in a diagonal pattern. With typical images this technique yields long runs of zeros, making the output very suitable for run length encoding (another example of manipulating the data to make best use of the available tools). The zero runs, and the remaining non-zero DCT coefficients are grouped to form *descriptors*, units that are particularly suited to entropy encoding. These descriptors are in turn encoded by an entropy encoder (usually a Huffman encoder) according to their frequency of occurrence. Finally, the resulting data stream is combined with various markers to indicate data types and boundaries; these markers also serve to aid re-synchronization at the receiver in the event of data errors.

Figure 5.6-5. Before (top) and after (bottom) JPEG compression. The lower image has been compressed by about 22:1.

At the receiver, the data stream is parsed to separate the different types of data. The entropy encoded data from each pixel block is decoded into descriptors which, together with the known descriptor definitions, allows the block of ac DCT coefficients to be re-assembled. Separately, the stream of data representing the dc coefficients is decoded first by the entropy decoder, then by the predictive decoder, and the appropriate dc coefficient is associated with each group of ac coefficients.

The next step is known as de-quantization. This is a misnomer as the process does *not* undo the effects of quantization. It does replace each quantized coefficient with the appropriate reconstruction value, taking into account the quantizer used for each individual coefficient.

All the coefficients of the block are now back in the correct range, and the inverse DCT transform is applied to generate a block of received pixel values. These are re-ordered into the correct scan or file format, along with any other color components, and the process is complete.

Figure 5.6-5 shows examples of fairly aggressive JPEG compression on the image known as *Boats*. The compression ratio is about 22:1, corresponding to about 0.36 bits per pixel.

On the one hand, it is easy to see the results of compression. There is loss of detail and very evident blocking artifacts in the sky. This large compression ratio is chosen to make the effects easily visible in the book. Nevertheless, the results are not all that bad. When the book is held at arm's length, the compression errors are not very evident.

In the enlarged views in Figure 5.6-6 the mechanism of JPEG can be seen clearly. The small blocks are pixels grouped into blocks of 8 × 8. It is easy to see that some blocks were coded with the DC coefficient only—they appear completely flat. In other blocks pat-

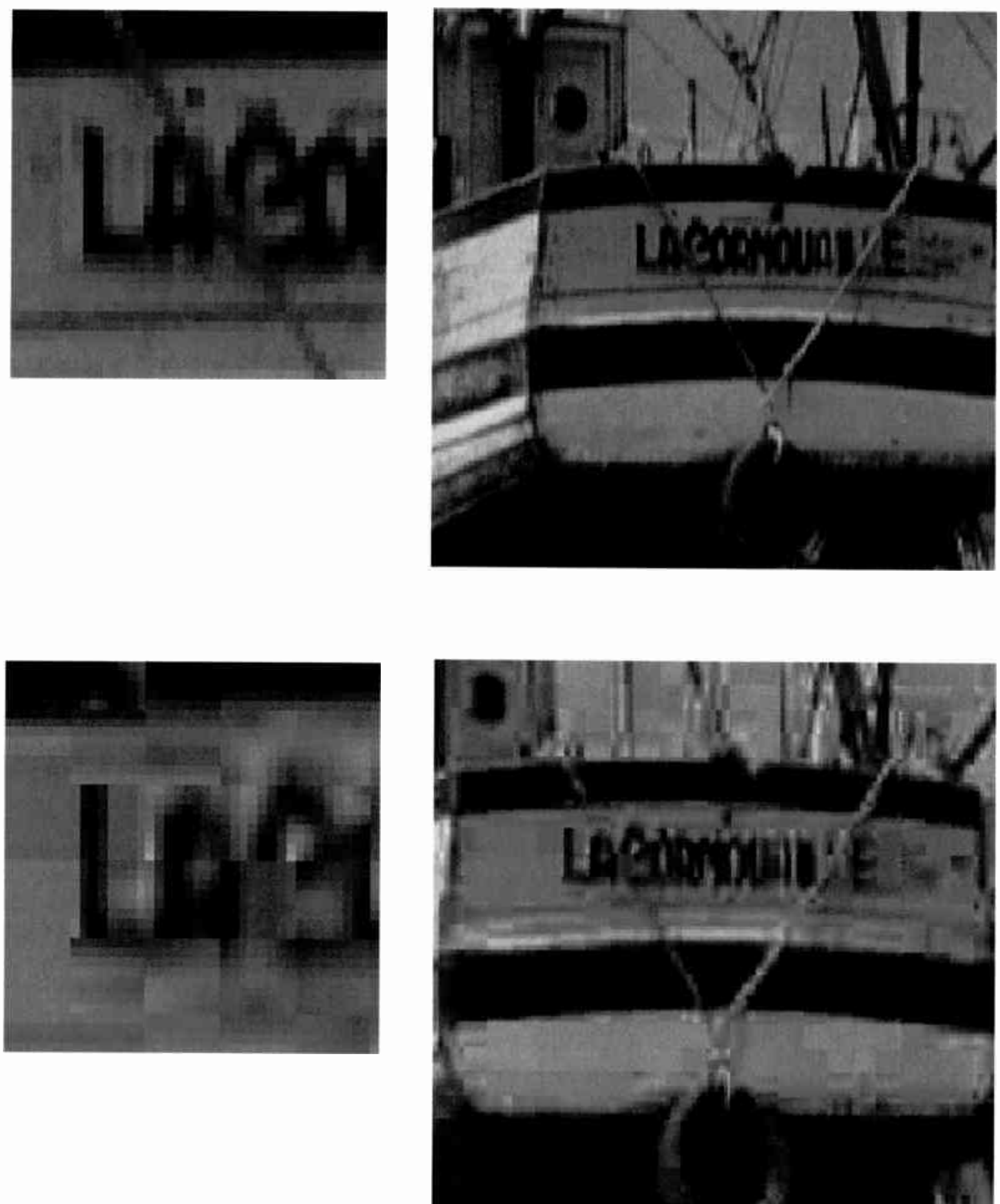


Figure 5.6-6. Enlarged detail from the images of Figure 5.6-5.

terns resembling the low order basis functions are clearly visible.

The previous description addresses the fundamentals of baseline JPEG. The standard includes many optional extensions, but these are beyond the scope of this chapter. The extension most important to television engineers is Motion JPEG. This is not part of the standard but, prior to the arrival of MPEG, it was the most powerful tool for compressing motion sequences. The concept is simple—each frame of an image sequence is coded as a JPEG image, and the frames are transmitted or stored sequentially. This is the basis of most non-linear editors, and most compressed video disk stores, today.

Unfortunately, because Motion JPEG is not properly standardized, virtually every implementation is proprietary and different. Usually image exchange must be performed by decoding in one equipment and re-coding in another. This is inefficient, but it is reality today. In most applications motion JPEG will be replaced by MPEG in the near future, and compressed interchange should be possible.

Motion Compensation

JPEG uses the concept of spatial correlation within a continuous tone to compress the image data. JPEG compression can be applied to a sequence of images, compressing each individually, with the technique known as Motion JPEG. In Motion JPEG, it is clear that the compression efficiency is a function of each individual image, and of that only. Each image in the sequence is compressed according to the same rules; it makes no difference whether the images are totally unrelated, similar in content, or even identical. Motion JPEG takes no advantage of any correlation between successive images.

However, it is easy to see that there should be considerable benefit available here. A sequence of images representing a moving scene is a set of temporal samples. In a typical scene there will be a great deal of similarity between nearby images of the same sequence. In other words, it should be possible to predict one from the other with some degree of success. As in the spatial domain, a compression system can use this property to separate and eliminate redundant data. The analysis of an image sequence to make predictions is called *motion estimation,* and the process of using these predictions to construct images is known as *motion compensation.*

The technique used is to find blocks of pixels, usually 16 × 16, that are identical or very similar in two images. First, one image is encoded by a spatial technique such as JPEG and transmitted to the decoder where it is decoded and displayed. This image will be stored at both the encoder and decoder, and will become the reference image. The second image is examined one block of pixels at a time. The motion estimator searches the reference image for a pixel block that is identical, or very similar to, the block being coded. Depending on the result of this search, various coding methods are available.

If an identical block is found in the same location (the picture at this location is unchanged), the block can be *skipped;* one simple code will instruct the decoder to use the block from the same location in the reference image. If an identical block is found in a different location, the decoder will have to be told where to find the block in the reference image. In this case a *motion vector* is transmitted, a two-dimensional value representing the horizontal and vertical offsets of the block found by the motion estimator.

The motion estimator may not find an identical block; instead it will return the motion vector pointing to the most similar block it was able to find. In this case, the encoder will subtract this block from the block being encoder to arrive at a block of residual values. This block of residuals may be spatially encoded and transmitted with the motion vector.

Static backgrounds and moving backgrounds provide a simple visualization of how motion vectors may be used to identify correlation between images in a sequence. However, it is evident that there is nothing special about the background—this technique can be used on any part of the image provided to find the same object, or part of an object, in two images of the sequence. The simplest technique, *block matching,* is to treat each image block separately, and search for an identical or similar block in the reference image.

The difficult part is finding the matching block due to the lack of adequate temporal sampling. If the temporal sampling obeyed Nyquist, tracking an object from one sample to the next would be a very easy task. Nyquist sampling implies the simplest possible path that connects an array of samples. It has already been established that the temporal sampling rate of any common imaging system is much less than that required by Nyquist rate. This fact leads to a simple conclusion—the position of an object in one image of the sequence provides no clue where it will be in the next.

Full search block matching tests every possible block within a defined search range against the block it is desired to match. The technique is accurate and exhaustive—if there is a match within the search range, this technique will find it. It is, however, computationally demanding.

MPEG

The Moving Pictures Experts' Group (MPEG), like JPEG, is a committee formed under the Joint Technical Committee of ISO and IEC. The MPEG-1 goal was to record video and stereo 48 kHz audio on a CD at the standard CD data rate of about 1.5 Mb/s. For this reason MPEG-1 is confined to a maximum image size of 352 × 288 for 50 Hz systems, and 352 × 240 for 60 Hz systems. This format, in its two variants, is known as *common intermediate format* (CIF), and is the image size used by many non-broadcast quality compression systems. Note that MPEG-1 has no tools for interlaced video.

MPEG-1 was "frozen" (subsequent changes were editorial only) in 1991. In the same year the MPEG-2 process was started, eventually becoming a standard in 1995. The initial goals were simple: to accommodate broadcast quality (standard definition) video and to accommodate interlace. In many ways, MPEG-2 represents the coming of age of MPEG. The greater flexibility of MPEG-2, combined with the increased availability of large scale integrated circuits, meant that MPEG-2 could be applied to a vast number of applications.

MPEG-1 is vastly more complex than JPEG, both in specification and implementation. Nevertheless, it is fair to look at MPEG-1 and characterize it as "slightly modified JPEG plus temporal compression and rate control." There are improvements on JPEG, but the changes are minor and based upon experience of an excellent system. This was not a given—at the beginning of MPEG many alternative spatial compression algorithms were tried, but the JPEG approach proved still to be superior. Similarly, MPEG-2 is based very heavily on MPEG-1. It is a measure of the quality of the work performed by these committees that each standard provided a solid basis for extension to the next.

MPEG-2 provides for coding of larger images and for interlaced video. It also adds layered coding schemes and a system layer, permitting multiple program streams (each with video(s), audio(s), and auxiliary services) to be carried in a single bitstream. The most fundamental addition, however, is a system of *profiles* and *layers*. MPEG-1 defined a single level of service; any MPEG-1 compliant decoder is required to decode any MPEG-1 bitstream. MPEG-2 permits such a wide variety of compression tools, and such a wide range of permissible bit rates, that the MPEG-1 approach is impractical. The set of permissible compression tools is defined by the profile, and the limits on picture size and rate, and bit rate, are defined by the level. Table 5.6-2 shows a few of the parameters that are constrained by particular profiles and levels. A particular use is defined by both sets of constraints, usually expressed as *profile@level*. The best known implementation is *Main Profile at Main Level*, usually abbreviated as *MP@ML*.

Some profiles of MPEG-2 permit encoding of 4:2:2 or 4:4:4 representations. The 4:2:2 profiles was added at the request of the professional television community, mainly for use in studios. Main profile met all the requirements for tools for this application, but supports only 4:2:0 coding. The 4:2:2 profile is defined by MPEG at Main Level, but its extension to High Level is being standardized by SMPTE for lightly compressed high definition signals in for studio applications, and for high definition program contribution and distribution links.

Perhaps the success of MPEG-2 is best highlighted by the demise of MPEG-3. This exercise was started with the objective of providing a compression system suitable for high definition television. It was soon abandoned when it became apparent that the versatility of MPEG-2 embraced this application with ease.

Table 5.6-2
The most common Profiles and Levels for MPEG-2.

Level	Profile: Simple	Main	4:2:2
High		1920H 1152V 60 Hz	1920H 1152V 60 Hz
High 1440		1440H 1152V 60 Hz	
Main	720H 576V 30 Hz	720H 576V 30 Hz	720H 512V/608V 30 Hz
Low		352H 288V 30 Hz	

Table entries represent maximum values.

MPEG-2 is, of course, the basis for the ATSC Standard for Digital Television now being implemented for both standard and high definition transmissions in the United States and other countries. The ATSC Standard employs MPEG-2 at *Main Profile at High Level (MP@HL)*.

It is worth noting the MPEG philosophy of standardization. MPEG describes various tools that may be used to perform compression, and gives some examples of how these might be implemented. The standard defines the syntax of a compliant bitstream and the ways in which a decoder must interpret valid bitstreams (those that conform to the defined syntax).

The really important thing here is the omission. MPEG does *not* define the encoder. A valid encoder is any device, implemented in hardware and/or software that produces a syntactically correct bitstream, resulting in the desired output if the bitstream is fed to a compliant decoder. This is a powerful approach, as no restrictions (other than intellectual property rights) apply to techniques or technologies used in the encoder. Any process, any algorithm, may be used in the encoder provided the output can eventually be expressed by some combination of the permitted syntactical elements. Given this, no matter what advances are made in encoder technology, any compliant decoder will produce the correct output.

MPEG-1 and MPEG-2 differ from JPEG in two major aspects. The first big difference is that both permit temporal compression as well as spatial compression. The second difference follows from this—temporal compression requires three-dimensional analysis of the image sequence, and all known methods of such analysis are computationally much more demanding than two-dimensional analysis. Motion estimation is the way we approach three-dimensional analysis today and practical implementations are limited by available computational resources. MPEG-1 and MPEG-2 are both designed as asymmetric systems; the complexity of the encoder is very much higher than that of the decoder. They are, therefore, best suited to applications where a small number of encoders are used to create bitstreams that will be used by a much larger number of decoders. Broadcasting in any form and large distribution CD-ROMs, are obviously appropriate applications.

The top level definition in MPEG is the *sequence* of pictures. A sequence is of arbitrary length and can represent a video clip, a complete program item, or a concatenation of programs.

Within the sequence, the next lower definition is the *group of pictures (GOP)*. The GOP is a very important concept in MPEG. A typical MPEG bitstream consists of a repeating GOP structure. In the simplest form of encoding, without temporal compression, the GOP can be a single picture, but in typical MPEG applications the GOP will be a comprised of pictures coded in three different ways, and arranged in a repetitive structure most commonly between 10 and 30 pictures long.

Below the GOP, the next step in the hierarchy is the *picture*, or *frame*, and one important point must be made. MPEG-1 includes no concept of interlace; every picture or frame contains all the lines that make up a complete image in the sequence. This is not the same as saying that MPEG-1 cannot encode interlace. The two fields of an interlaced frame can be combined as one frame, but MPEG-1 makes no concession to this practice. There are no special tools for interlace, and no means of avoiding the coding inefficiencies that may result from interlace (in an interlaced system the frame will consist of lines that were captured at different times). MPEG-2 offers a number of tools for coding interlaced pictures, and the combination is extremely efficient.

Continuing down the hierarchy are the *slice* and the *macroblock*. A macroblock contains all the information required for an area of the picture representing 16 × 16 luminance pixels. Macroblocks are numbered in scan order (top left to bottom right). In MPEG-1, a slice is any number of sequential macroblocks; MPEG-2 restricts this to a single horizontal row, or less, of macroblocks. The chief significance of the slice is that it is encoded without any reference to any other slice; this means that if data is lost or corrupted decoding and recovery can usually commence at the beginning of the next slice.

The prefix *intra* means *inside* or *within*, and an *intra frame* or *I-frame* is a frame that is encoded using only information from within that frame. In other words, a frame that is encoded spatially with no information from any other frame—no temporal compression. Coding of an I-frame is similar, but not identical, to coding of an image in JPEG, or of a single frame in motion JPEG.

Non-intra frames use information from outside the current frame; from frames that have already been encoded. With non-intra frames motion compensated information, as described in the previous section will be used for a macroblock where this results in less data than directly (intra) coding the macroblock.

There are two types of non-intra frames, *predicted frames (P-frames)* and *bidirectional frames* or *(B-frames)*. These will be described with reference to the series of images shown in Figure 5.6-7.

The first image *X* shows a background, a tree and a car. The third image *Z* shows the same scene somewhat later. The camera has panned to the right, causing all static objects to move left within the frame, but the car has moved to the right, and now obscures part of the tree. The second image *Y* shows a point in time somewhere between *X* and *Z*.

Image *X* will be encoded as an I-frame (it is spatially encoded without reference to any data outside that frame). When the frame has been processed, the encoded data is sent to the decoder where it is decoded and the reconstructed image stored in memory. At the same time, the encoded data is decoded at the encoder to provide a reconstructed version of *X* identical with that stored at the decoder. This is the reference picture.

Image *Z* will be encoded as a *P-frame*. For each macroblock in *Z* the encoder will search for a matching macroblock in the reference picture. The objective is to find a motion vector that links the macroblock to an identical or very similar macroblock in the reference. As described in the previous section, residuals will have to be considered if the macroblocks are not identical.

Assuming that the motion detector is good, and has an adequate search range, we should be able to encode a large proportion of image *Z* with respect to *X*. Notable exceptions are the area behind the car in *X*, and the strip of image at the right-hand side of *Z* that was revealed by the camera pan—this information just did not exist in picture *X*.

This gives a clue as to how to improve the efficiency of temporal encoding. In the image sequence shown *X* is the earliest in time, followed by *Y*, followed by *Z*. The coding of *Z* has been described, but as yet the encoding of *Y* has not been discussed. Suppose picture *Y* is stored away until after *Z* is encoded. Both *X* and *Z* can be used as reference frames while encoding *Y*. In this idealized case it should be easy to find a motion vector for almost every macroblock in *Y*. This type of encoding is known as *bidirectional encoding*, and frame *Y* is said to have been encoded as a *B-frame*.

A typical MPEG GOP is shown in Figure 5.6-8. The pictures are now categorized in a slightly different way. I-frames and P-frames are called *anchor frames*, because they will be used as references when coding other frames using motion compensation. B-frames, however, are not anchor frames, as they are never used as a reference.

The GOP shown starts with an I-frame. Clearly, it is essential to *code* an I-frame first to start the sequence; if no previous information has been received there is no possible reference for motion estimation. It is possible to have a number of B-frames precede the I-frame provided these are encoded and transmitted after the I-frame. The first P-frame is coded using the previous I-frame as a reference for temporal encoding. Each subsequent P-frame uses the previous P-frame as its reference. (This shows an important point: errors in P-frames can propagate as the P-frame becomes the reference for other frames.) B-frames are coded using the previous anchor (I or P) frame as a reference for forward prediction, and the following I or P frame for

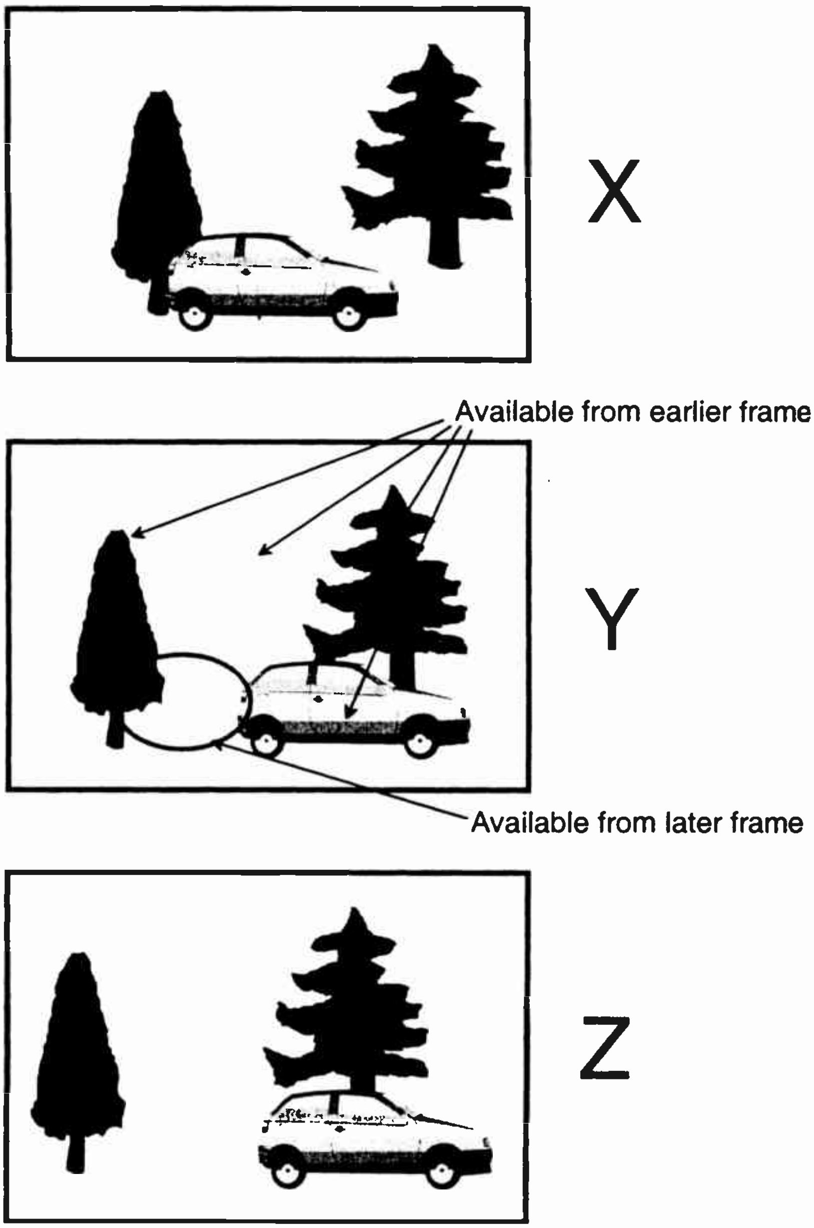


Figure 5.6-7. Bi-directional motion estimation.

backward prediction. B-frames are never used as a reference for prediction.

In a *closed GOP*, as shown in Figure 5.6-8, all predictions take place within the group of pictures. Many MPEG practitioners prefer this approach, but there are examples of *open GOP* structures (I-B-I-B-I). Such structures can be quite efficient, but there is no point at which the bitstream can be separated, every frame boundary has predictions that cross it.

Note that B-frames can only be decoded if both the preceding and succeeding anchor frames have been sent to the decoder. Figure 5.6-8 shows the GOP in *display order*, but to enable decoding the frames are actually transmitted in a different order, as shown in Figure 5.6-9. It is important to note that this re-ordering, essential when B-frames are used, adds substantially to the delay of the system.

It can be seen from the preceding discussion that the coding decisions for a macroblock depend upon the type of frame. However, before considering this, we need to look at the structure of a macroblock and the various possible ways of coding it.

MPEG-1 and MPEG-2 Main Profile use Y, C_B, C_R coding and a 4:2:0 structure of color information. This means that the luminance is coded every pixel, but the color difference information is filtered to half the luminance resolution, both horizontally and vertically. Therefore, an image area represented by a block of

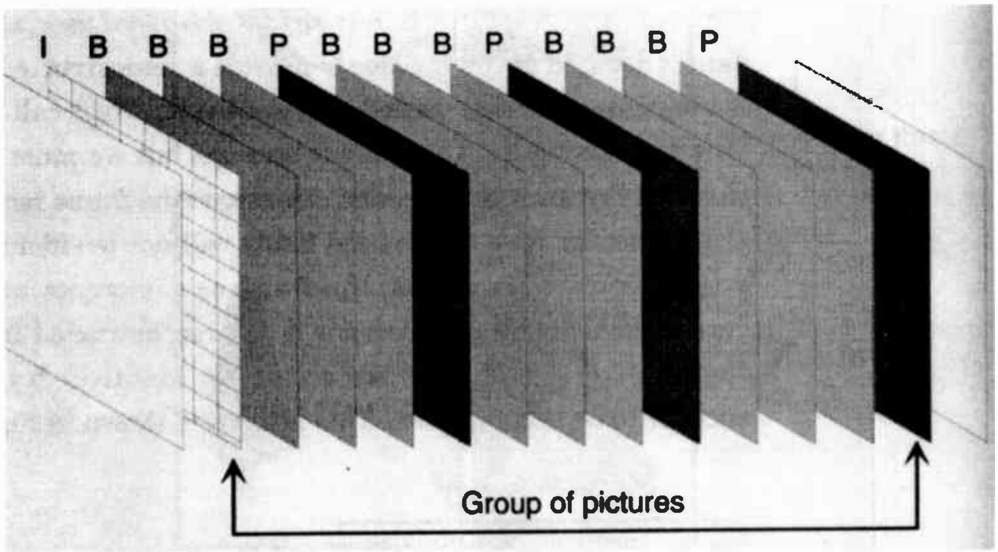


Figure 5.6-8. A typical group of pictures in display order.

luminance pixels 16 × 16 requires only 8 × 8 for C_B, and 8 × 8 for C_R. When using 8 × 8 blocks for all DCT coding, the macroblock consists of four blocks of luminance samples and one block each of C_B and C_R samples. (Note that 4:2:0 coding is not identical in MPEG-1 and MPEG-2; the location of the C_B and C_R pixels is defined differently.)

When a macroblock is intra coded, the procedure for each block of samples is almost exactly the same as described for JPEG but there are a few significant differences. Unlike JPEG, MPEG does define default quantization tables, one is used for intra coding, the other used to code any residuals when inter coding. Either or both quantization tables may be changed by specifying new tables in the sequence header, otherwise the decoder will assume the defaults.

MPEG scales the quantization by means of a *quantization scale factor*, sometimes known as *M-Quant*.

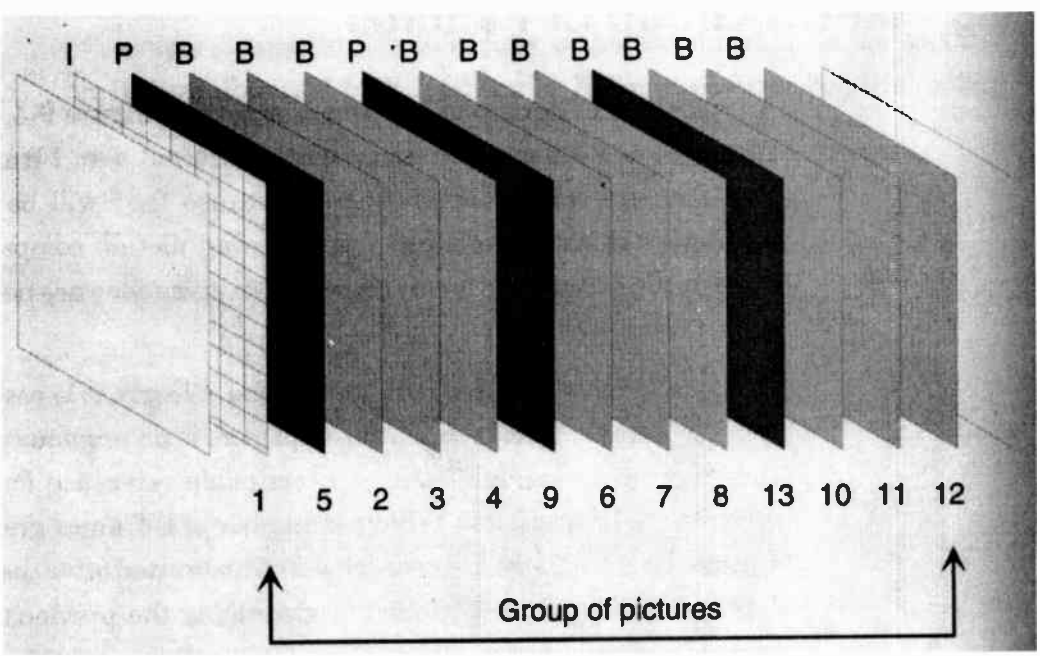


Figure 5.6-9. The same group of pictures as shown in Figure 5.6-8, but pictured in coding (and transmission) order. The numbers below the frames represent display order.

This value is set explicitly for every slice, but may be changed at the macroblock level if required. Each macroblock is assumed to use the same value as the preceding one, unless a new value is set in the macroblock header. M-Quant values range from 1 to 31. The ac coefficients are quantized by multiplying by eight, then dividing by the appropriate value from the quantization table, and by M-Quant. An M-Quant value of one will decrease the severity of quantization by a factor of eight (higher quality); a value of 31 will increase the severity by a factor of four (approximately), yielding lower picture quality.

As in JPEG, the dc coefficients are extracted and coded predictively. The ac coefficients are quantized as previously discussed, then scanned and variable length encoded almost exactly as described for JPEG.

In a P or B-frame, it may be necessary to intra-code a macroblock just as with an I-frame. If there is not a reasonable match in the motion estimator, or if the best match requires transmission of substantial residuals, it will revert to intra-coding of that macroblock. In MPEG, only the luminance samples are used in motion estimation, but the resulting motion vector is used in coding both luminance and color difference blocks.

When the motion estimator has found the best match, each block within the macroblock will be treated separately. Each block being encoded will be subtracted (pixel-by-pixel) from the corresponding block in the matching macroblock, leaving an array of error values. This block of error values (residuals) will be encoded by DCT transform and quantization as in intra-coding, except that a flat inter-quantization matrix is used. This process is applied to the six blocks that make up the macroblock (four luminance plus two color difference). If the motion estimation is good, some blocks will likely produce small residuals that quantize to zero throughout, and these blocks do not need to be encoded at all and can be skipped. When all blocks have been either encoded or skipped, a code is chosen to represent the pattern of blocks that have actually been encoded. If no blocks need encoding, the entire macroblock may be skipped.

Motion vectors also need to be coded. For the first macroblock in a slice, the motion vectors must be sent in full. For the rest of the slice, the vectors are predictively coded; in other words the vectors are assumed to be the same as for the preceding macroblock and differential values are sent for correction if required. Both motion vectors and differentials are variable length encoded.

When encoding a P-frame, there is a choice of coding methods. If it has been determined that the entire macroblock can be skipped, the decision is easy, simply write a *skip* code into the macroblock header and proceed to the next macroblock. Otherwise, total the number of bits that will be required to transmit the motion vectors, the code showing which blocks of residuals have been encoded, and the coded residuals, and compare this total with the number of bits required to transmit the macroblock intra-coded. If the motion compensated coding uses fewer bits, that is what is transmitted. Otherwise the macroblock is marked as intra-coded and that data is transmitted. (In this case, the next macroblock assumes a value of zero for the preceding motion vectors when predictively encoding its motion vectors.)

When encoding a macroblock in a B-frame, there are a few more calculations to perform. As with a P-frame, the number of bits for intra-coding is calculated, and the number for inter-coding with forward prediction. Because this is a bidirectionally encoded frame, two more operations are performed.

The motion estimator will also find a backwards prediction, and the encoder must go through exactly the same process as previously described to determine a total number of bits for inter-coding with backward prediction.

Finally, the encoder may assess a fourth method of encoding that is allowed for bidirectionally encoded macroblocks, interpolated prediction. For regular motion prediction, the macroblock is used at the end of the motion vector as a predictor. This is then subtracted from the macroblock being coded. In interpolated prediction the mean (on a pixel-by-pixel basis) of the macroblock at the end of the forward motion vector and the macroblock at the end of the backward motion vector are considered. This mean is used as the predictor. It is subtracted from the macroblock being encoded to form the residuals. So, if "C" is the current macroblock, "F" is at the end of the forward motion vector, "B" is at the end of the backward vector, and "R" is the macroblock of residuals:

$$R_{i,j} = C_{i,j} - \left(\frac{F_{i,j} + B_{i,j}}{2}\right)$$

where i and j range from 0 to 15 to cover all 256 samples.

Note that this is an unweighted average, irrespective of the position of the B-frame with respect to the surrounding anchor frames. The mechanism is, apparently, useful when encoding noisy pictures.

Any one of the inter-codings may allow the macroblock to be skipped, although in this case we need to specify which of the prediction modes provided that condition. Otherwise, the bit totals for the three prediction modes are compared. If the best of these needs fewer bits than intra-coding then that mode is used, otherwise the intra-coding is transmitted.

Perhaps the most difficult concept associated with MPEG is rate control. Consider the GOP structure. Even if there were no issue in distributing bits throughout a picture, the different coding methods create a problem in rate control. Intra-coding is necessary, otherwise there is no place to start when creating video in the decoder. Obviously inter-coding must be more efficient, or compression system designers would not go to the enormous difficulty and expense of implementing it. A P-frame will typically use fewer bits than an I-frame, and a B-frame should use fewer still. This means that over a transmission system that uses a constant or nearly constant bit rate, I-frames will take longer to transmit than P-frames, which will take longer than B-frames. However, at the encoder the unencoded frames are arriving regularly every 1/30 of a second, and the decoder must send a frame to the display every 1/30 of a second. In turn, this means that the delay from encoder input to decoder output must be constant. This can only be achieved with a buffer at each end.

Actually, the situation is far more complex than this. A good MPEG encoder has to apportion bits according to the image type, the overall complexity of each image, and the differing complexities within each image. It would be reasonable to expect a formidable array of tools to help control bit allocation. In fact, there is one tool—the quantization scale factor, or M-Quant. Everything else will minimize the number of bits for each macroblock for a given M-Quant, but nothing else provides any type of control as to where the bits are used.

The encoder produces a variable rate stream of bits that go into a buffer. The buffer is emptied by a constant transmission rate. If the encoder produces bits too slowly, the buffer will empty, and there will be nothing to transmit. If the rate from the encoder is too high the buffer will overflow and data will be lost. To try to stay within these two bounds the control system must feed a measure of buffer fullness to the magic box called the rate controller, which responds by varying the quantization scale factor, affecting quantization of the DCT coefficients of blocks that have not yet been coded!

In the real world delivery of useful program material requires audio as well as video, and may require additional data streams such as captioning. Clearly it is not a good idea to send all the video, then start the audio, so the bitstreams need to be *multiplexed* together. In some situations it may be desirable to send several different programs, each with its associated video, audio, etc. in the same bitstream. Because continuous information (such as audio) is being transmitted in a discontinuous manner (packets), there is a need for timing and control to ensure that all the inputs can be re-assembled correctly. All of these issues are handled by the MPEG systems layer that defines two types of bitstream, the *program stream* and the *transport stream*. Program streams are best suited to variable rate, low error, applications such as CD replay. Transport streams are intended mainly for fixed rate applications, and use short, fixed length, packets of data that are very suitable for forward error correction. This is the stream type used by the ATSC Digital Television Standard.

OTHER COMPRESSION SYSTEMS

JPEG and MPEG together represent most of the television broadcast applications of compression today. In particular, MPEG-2 is used in DVD, and in the ATSC Digital Television Standard. This section mentions briefly some of the other systems suitable for video.

The DV compression standard has become very popular, particularly for news applications. The system was designed to be very economical so as to be suitable for consumer applications; the same chip performs compression and decompression. DV is an intra-field system using the DCT transform, followed by quantization and lossless compression. Incoming data is shuffled to distribute complex blocks evenly, and there is a mechanism that allows left-over bits to be used by more complex blocks.

Basic DV encodes a 4:1:1 color space with approximately 5:1 compression to give a 25 Mb/s bitstream. A professional varient uses two chips and encodes 4:2:2 yielding a 50 Mb/s bitstream. In this mode the compression ratio is only 3.3:1, and the system is very close to transparent.

SX compression is an implementation of MPEG 4:2:2@ML using an open GOP with an I-B-I-B structure and a bit rate of about 18 Mb/s. The open GOP structure makes insert editing complex.

Wavelets have been used in some television applications. Some suggest that wavelets are well suited to low bite rate applications, but to date DCT has proved superior for near-transparent broadcast applications.

Finally, a word about MPEG-4. Initially intended just to address very low bit rate applications, the work has expanded to cover a wide range of bit rates, and a system permitting separate transmission of picture elements that may be composited in the decoder. It is likely that this system will have a major effect on computer applications using internet communications. However, the large number of decoders now being deployed into the consumer market suggests that MPEG-2 will dominate this market for some time to come.

5.7
VIDEO RECORDING PRINCIPLES

STEVE EPSTEIN
BROADCAST ENGINEERING AND *VIDEO SYSTEMS* MAGAZINES
OVERLAND PARK, KS

INTRODUCTION

Before the invention of the videotape recorder (VTR), film provided television's only video storage method. Images (and sound) were recorded using standard film equipment, either in the studio or on location. Specialized projectors and television cameras were used to convert filmed images into video. Images were also recorded onto film directly off special, high quality television monitors for later playback. Devices such as these, called *kinescopes*, were used to time shift network feeds to the west coast. Unfortunately, exposed film had to be developed before it could be used, adding considerable time to the procedure. Film could be used for recording only once, and the networks required a considerable amount of film to time shift an entire day's broadcast. By 1954, American TV operations used more raw film than all of the Hollywood studios combined. A rerecordable medium that allowed for immediate playback was needed. As it turned out, that medium was videotape.

Immediately prior to the 1956 NAB convention, on April 14, Ampex introduced the first practical VTR to a meeting of CBS station executives. Later that year, on November 30, the VTR was first used on-air for the CBS airing of *Douglas Edwards and the News*. In the forty years since, the VTR has gone from technological breakthrough to commodity.

Within the broadcast and professional community, the VTR remains a resounding success, but in this age of computer based systems, hard disk drives, developed and used throughout the computer industry are finding popularity as video recording devices. In addition to computer based hard drive systems, specialized disk recorders are also finding use. Although many of these video disk recorders (VDRs) record digital bits, a handful are analog devices. For the most part, this chapter examines professional video recording systems, however, some information relative to popular consumer devices is also included.

FUNDAMENTALS

Nearly all video recording today is done using magnetic media, either tape or disk. Some optical or magnetooptical devices are also used; many in specialized applications including archive libraries, kiosks and short form programming such as CD-ROM or DVD (Digital Versatile Disk) applications. DVD is just becoming available at the time of this writing, and its level of acceptance into the marketplace (both consumer and professional) remains to be seen. DVD offers considerable longer play times than CD-ROM, especially on the larger disk sizes.

For any video recording method, some fundamental design choices must be made. The first is whether the recording is to be analog or digital. For analog recordings, there are bandwidth issues to be decided, whereas for digital, the choice is the number of bits to record. Most popular recording systems use 8 bits. The second choice is whether to record composite or component video. Early recorders were composite, while much of today's equipment records the video components individually. One additional choice is whether to record on tape or disk. Videotape, in terms of dollars per minute is by far the most cost effective, however, videotape machines can be expensive, both to purchase and maintain. Disk drives require only minimal maintenance, but typically do not provide the long record times offered by tape. Unlike videotape, if a disk drive fails, the media cannot be easily moved to another machine. Another consideration of disk storage is the type of accessibility. Some of this is dependant on the VDR chosen, but it boils down to whether the video is stored, handled and output as video, or if it is dealt with in terms of a computer file. A few recorders can handle either method gracefully. Both systems have their merits. The advantages and disadvantages many times come down to the intended application within the facility.

Several criteria can be used to categorize videotape machines. These include:

- *Physical tape size:* Tape widths range from 1/4″ (6.35mm) up to 2″, Lengths, rather than being expressed in terms of feet or inches, are expressed in minutes or hours, with 2–3 hours being the typical maximum
- *Tape style:* These include open reel and cassette. In the professional arena, only the 2″ and 1″ formats were open reel machines, all others have used cassettes. Today's machines are almost exclusively cassette based
- *Scanning method:* Transverse and helical scan are the only successful types. The 2″ quadruplex format utilizes transverse scanning, all others have been helical scan (see Figure 5.7-1)
- *Recording format:* Formats include composite (direct-color), Y/C (color-under) and component (Y,

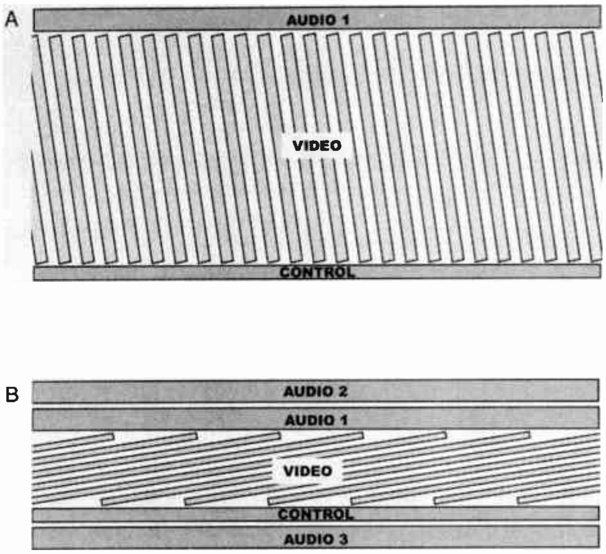


Figure 5.7-1. Scanning system used for videotape recording, (a) transverse scan, (b) helical scan.

R-Y, B-Y) methods in both analog and digital designs
- *Input/output connections:* Most machines provide input and output connections for analog composite video and two channels of balanced audio. Some machines provide analog component video I/O. In addition, some machines provide for up to four analog and/or digital balanced audio connections. S-video (separate luminance and chrominance on a single 4-pin mini-DIN connector) and dubbing connections are common on many of todays VTRs. Digital video, either component or composite is available on nearly all of the VTRs designed to record video digitally.

Video disk recorders have similar criteria including:

- *Maximum record time:* This is based on the amount of storage within the device. Some devices can be expanded through the use of external drives. Storage time may be for video and two audio channels, or for video only
- *Compression requirements:* Compression may or may not be required depending on internal bandwidth issues. Some units may have selectable compression rates, but require all clips played in a sequence to be compressed to the same level. Other units may offer the ability to playback various compression levels randomly
- *I/O connections:* Units may have video I/O, LAN connections, or disk drive connections such as SCSI. Some units provide all of these and more. VDRs may be video only, or offer video and audio recording. Some units are capable of handling multiple I/O streams simultaneously
- *Redundancy/backup:* Magnetically recorded media typically cannot be separated from the complete drive assembly, therefore, some level of redundancy (RAID) or backup mechanism may be required for critical applications
- *Random access:* One of the primary features of VDRs is their ability to randomly access clips. Not all drives are created equally. A few units can truly play any frame after any other frame, however, some require sequences to be prerendered for real time playback.

The basics of today's VTRs are similar to the machines produced in the mid 60's. The 2″ VTRs are relatively straightforward and easy to understand, and having a basic knowledge of these formats makes it much simpler to grasp the concepts behind the newer machines. Because of this, many of the descriptions are based on the 2″ machines.

VTR basic building blocks include:

- *Transport and related control systems.* The transport is an electromechanical assembly that provides a precise path through which the tape moves. Also included are the various motors and guides used to move and position the tape. Magnetic tape transports require mechanical alignments to be maintained precisely. Tape must be capable of moving at any speed and in the forward or reverse directions without experiencing any edge damage. This is especially true in editing machines. Control of the early VTRs was accomplished using simple relay circuits; today's units rely heavily on microprocessor based systems
- *Servos.* Used to control the head drum assembly, capstan and in many units, the reel motors
- *Video and audio circuitry.* Basic video circuits consist of input, modulator, demodulator and output stages. Some newer machines include bit rate reduction (compression) circuitry as well as A/D, D/A convertors and/or composite to component encoder/decoders. Audio in most VTRs is comparable to professional audio recorders
- In many cases, additional circuitry for editing or other special functions such as dynamic tracking.

Tape Transport and Control

Fundamental to any videotape machine is the mechanical transport assembly. The transport provides the mechanical interface between the machine's electronics and the information recorded on the tape. Depending on the format, the transport will include the loading mechanism and reel hubs, tape guides and sensors, the drum assembly and the capstan. Recording/playback heads include the erase head, audio heads, a control track head and the various video heads within the drum assembly. Additionally, the transports of professional editing machines typically include time code heads, audio confidence playback heads, audio-only erase heads and flying erase heads mounted in the drum assembly. Flying erase heads are used during insert edits and at the beginning of assemble edits, until sufficient time has passed to allow the full track head to begin erasing the tape (if necessary).

Constant torque motors linked to the reel hubs prevent slack tape and minimize tape stretch caused by excessive tension. Brakes are engaged when tape is stopped to prevent spillage. On many units, the brakes are also used to slow the tape during movement, however, more sophisticated machines rely on eddy currents within the motors to provide gentler tape deceleration.

Precision rotating and fixed guides are used to route the tape around, over and through the various assemblies of the transport. Some machines also make use of air (either pressure or vacuum) to position, handle or buffer the tape. The 2" machines used a vacuum guide to control and position the tape at the head drum assembly. Several formats have used vacuum capstans to precisely control tape movement without a pinch roller. Pressurized air escaping through small holes in fixed guide assemblies provides an *air bearing* reducing both tape and guide wear. Loading and unloading of tape transports has also been accomplished quickly and accurately using air pressure, vacuum or both.

Head drum assemblies contain one or more video heads. For professional level recording, relative tip to tape movement is typically on the order of 1,000 ips (inches per second). A capstan/pinch roller combination or a vacuum capstan is used to precisely control tape movement. A tracking control is used to electronically change the mechanical timing relationship between control track pulses and the recorded video tracks. Non-standard recordings may have these pulses delayed or advanced with reference to the video tracks. Either way, the RF signal recovered from the tape is less than ideal, as the head is not centered on the recorded tracks. Proper adjustment of the tracking control centers the heads on the recorded tracks and allows maximum RF to be recovered from the tape.

Control circuitry provides the proper signals to the various motors involved. Early control systems were rather primitive when compared to today's machines, which rely almost exclusively on microprocessor based control systems. Relays or transistor drivers provide the necessary power to the motor. Modes include what has become the standard transport controls: Stop, Play, Record, Fast Forward and Rewind. Most professional machines have forward and reverse search modes which combine viewable pictures with rapid tape movement. Dynamic tracking, provides a playback mode in which the tape moves at a non-standard rate. Fields (2 per frame) are skipped or dropped as necessary. A ready mode on some machines provides a means to get the head drum assembly up to speed without rolling tape, allowing for shorter preroll times. In addition, an electronics to electronics (E–E) mode provides a path from the input circuitry to the output circuitry during stop mode.

Additional control circuitry is required based on the transport type. Open reel machines typically have tape in path detectors that consist of either spring loaded tension arms or optical sensors. Few provisions are made for threading other than brake release switches. Newer machines detect near end of tape based on reel pack diameters and can be instructed not to cross into the near end of tape zones, preventing accidental unthreading at tape ends. Machines capable of this type of detection generally offer an *unload* mode in which the tape is unloaded at a slow speed. Head drums can be damaged by wrinkled tape ends being pulled through at high speed. Ragged tape ends should be cut off, and tape ends should lay flat on the takeup and supply reels of open reel machines, reducing the possibility of damage to the video head assemblies.

Moving from the open reel to the cassette machines, early cassette transports were of the *top load* variety. Cassettes were inserted into a bin which was lowered into the deck. Most newer studio units are *front loaders* and like today's home units will take the tape from the operator once inserted. Many portable units designed for field use are of the push to load variety. Elevators of some type are required on nearly all videocassette formats, because tape must be removed from the shell and threaded through the transport mechanically. To accomplish this, guides must be behind the tape to pull it out of the shell. The elevator moves the cassette above the guides. Once the tape is past the guides, the cassette is lowered, positioning the tape between the guides and the transport. A door assembly is used to protect the videotape from contamination when the cassette is out of the machine. Many newer cassette formats have a clamshell door assembly that protects both sides of the 1nbtape. Reel locks, located inside of some cassettes, prevent the hub assemblies from rotating when the cassette is out of the machine. A pin assembly located in the transport disengages the brakes as the elevator positions the cassette within the transport.

Servo System

VTR servos include those used for the capstan and drum. *Capstan servos* are responsible for tape movement during most modes that provide viewable pictures. Drive oscillators typically have a wide range to provide the tape speeds required. Error correction signals are provided through a feedback loop that includes the control track head. The capstan servo must position the tape such that recorded tracks are physically located where they can be read by properly positioned video heads. The *drum servo* positions the drum assembly containing the video playback head such that it can properly playback the signal recorded on tape. Nearly all professional units allow the capstan and drum servos to be referenced and locked to external sync. Once a transport is locked to external sync, it is relatively easy to lock two transports for editing purposes. *Reel servos* are used in many newer VTRs, but some, especially small format machines, rely on constant torque motors. Felt clutches or brake bands on the reel tables provide takeup and holdback tension. Some high-end machines tie the reel servos to the motion of the capstan, assuming if the capstan is stationary, the tape is not moving. *Capstan motion*, along with pulse generators located in the reel tables can be used to calculate tape pack diameters precisely. These calculations, along with tape tension detectors provide sufficient information to gently and precisely handle tape in a manner that allows nearly instantaneous high speed acceleration and deceleration.

Video Signal Path

The electronic components used in VTRs as well as the actual circuitry have evolved considerably over the last 40 years. However, the basic building blocks are comparable. There are numerous variations depending on the format and actual recorded signal. Consider the task of recording a composite signal. On the record side, 1V pp composite video goes into an input amplifier and sync separator. Sync pulses are routed to the servos so the transport mechanisms can be properly phased and standard recordings can be made. Once the tape is in motion and the servos are locked, a record lockout circuit verifies that the tape can be recorded on. The erase heads are energized, erasing any previous recordings and preparing the tape for the new one. The video signal is sent through a preemphasis circuit and frequency modulated. Rotary transformers complete the circuit to the record amplifier and record heads. Many machines provide inputs and output connections at the record amplifiers and playback preamplifiers. In this manner, the modulated RF signal can be dubbed from one machine to another without having to be demodulated to baseband and then remodulated for the duplicate recording (dub).

Upon playback, the capstan and drum servos position the tape such that the reproduce heads are properly aligned with the recorded tracks. Relative movement of the tape and head assemblies causes an induced current in the heads proportional to the recorded information on the tape. Preamplifiers located close to the video heads, typically within the spinning drum, are used to increase signal levels, thereby minimizing noise. Once amplified, the reproduced signal is equalized to produce flat frequency response. Rotary transformers provide a signal path from the spinning heads to the stationary transport. A limiting amplifier and demodulator complete the playback signal path. Because the recorded signal is frequency modulated, amplitude variations in the reproduced signal do not affect the reproduced signal. Passing the signal through a limiting amp provides the demodulator with a constant level signal. In the demodulator, the carrier frequency is removed from the signal, and only baseband video remains. The baseband video is put through a deemphasis circuit to remove the frequency distortions which were added to improve the recording process. Finally, an output amplifier is used to set the signal level and provide sufficient power to drive the unit's outputs (see Figure 5.7-2).

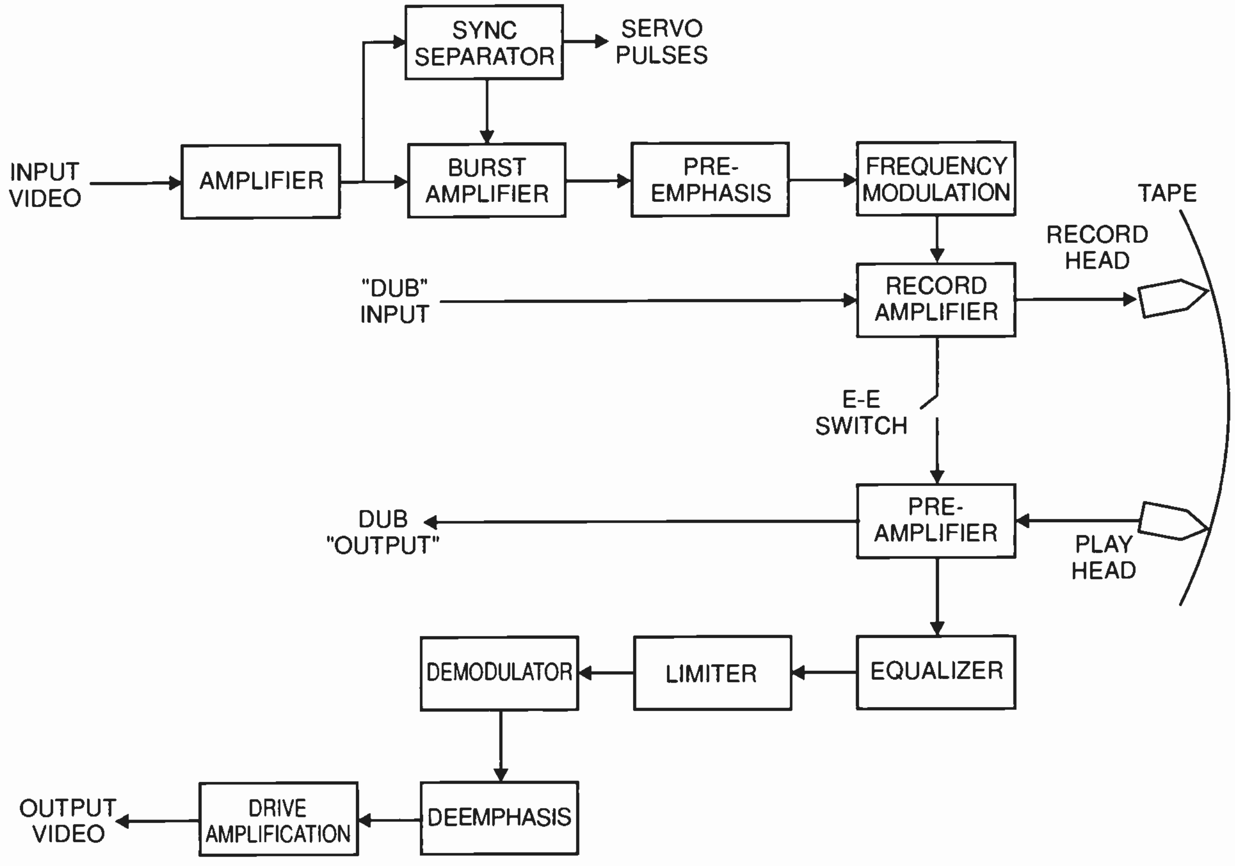


Figure 5.7-2. Block diagram of the video record and reproduce circuits used in a typical videotape machine. (Adapted from *NAB Engineering Handbook, Eighth Edition p. 880*)

Generally, the output amplifiers are considered the final stage in the playback circuitry. However, due to time base errors caused by both the playback and record processes, additional circuitry is required to stabilize and correct the signal to meet the requirements of RS-170(A). Time base correctors (TBCs) have become so common place that today complete units are available as plug-in cards for personal computers. Many videotape machines, typically those of the color-under variety, require minimal stabilizer circuitry to provide output signals suitable for display on a color monitor or television. Although these machines provide viewable pictures, a TBC is recommended if the video output is to be used for anything other than viewing.

Dropouts are fairly common on video recordings, and are caused by either airborne particulates or small sections of tape with poor response. Particulates that end up between the head and tape during record will prevent RF from being recorded, thus causing a permanent loss of signal. During the playback process, particulates will cause random and generally non-repeatable dropouts. Tape response problems are caused either by physical damage or a lack of oxide. Manufacturing techniques have improved considerably over the years and high quality tape is readily available. Today, airborne particulates are the primary cause of dropouts not caused by physical damage. Maintaining a clean environment is the best way to reduce dropouts of this nature. In analog recordings, dropout compensation is typically accomplished by repeating previous lines whenever a loss of RF occurs. For digital recordings, error correction and concealment techniques are used to prevent noticeable signal loss.

For the most part, the previous description applies to the composite recording process used in the 1″ type-C format. The 2″ quads used similar circuits, however, the video signal was segmented. Four separate heads were used to record and playback the video signal. For color-under recordings, including the 3/4″ U-matic, S-VHS and Hi8 formats, two parallel circuit paths are used, one for luminance and the other for chrominance. These units typically provide composite and Y/C inputs and outputs. Many decks also provide dubbing connections. Although the distribution of separate Y/C signals throughout a facility may be considered *component distribution*, the color-under recording process is not considered a component format. In the color-under process, composite input signals are fed through a low pass filter, limiting high frequency luminance components to approximately 3 MHz and separating the luminance signal from the chrominance signal. The chrominance subcarrier is bandpass limited to provide a 0.5 MHz wide double sideband signal. This signal is amplitude modulated onto a carrier near 650 kHz, which is added to the FM luminance carrier for recording on tape. The recording spectrum is shown in Figure 5.7-3.

At present, all color-under systems are helical scan machines. Normally two heads are used to record the combined signals, one for each field. Typically the tape wraps 180° around the head, allowing the heads

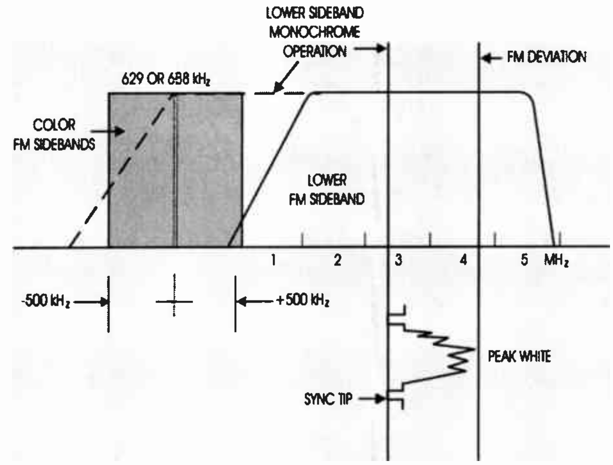


Figure 5.7-3. The recording spectrum used by video tape machines that use the color-under recording process. The AM chrominance subcarrier is added to the FM luminance carrier and recorded on tape. Part of the lower sideband of the FM luminance signal is removed to allow for the chrominance signal. (Source: Benson and Whitaker, *Television Audio Handbook p. 7.7*)

to be switched by a symmetrical square wave locked to incoming vertical sync or, during playback, generated by the control track signal. One advantage of the color-under process is that time base errors in the color signal are kept to a minimum, allowing the output signal to be displayed on a monitor or television with little or no correction circuitry. Disadvantages include high chroma noise and limited chroma and luminance resolution. Many of the smaller formats improve tape packing density by eliminating the guard bands between recorded tracks. This is accomplished by angling the head gap relative to the recorded track. Rather than being perpendicular to the track, the gap is up to ±15° from perpendicular. One head is angled one direction from perpendicular, and the other head is angled the other way. Because of this, guard bands are not needed as the off-axis response of adjacent tracks is minimized.

Like the color-under formats, the component formats separate luminance and chrominance. However, these formats take an additional step and separate the chrominance into two color difference signals, R-Y and B-Y. The two primary formats using this recording process are M-II from Panasonic and Sony Betacam (SP). These two formats are similar but not interchangeable. Although the same component signals are used for input and output (Y, R-Y and B-Y), signal levels are different. Care must be taken to adjust I/O levels when decks based on these different formats are interconnected. Most decks provide composite and component inputs and outputs as well as dubbing connectors. These units provide reasonably high quality encoders and decoders for composite to component signal conversion. Three parallel signal paths exist, the luminance path typically has a 4.2–4.5 MHz bandwidth, while the paths used for the color difference signals have only a 1.5 MHz bandwidth each. On these machines, one head records the FM luminance signal on a single track per field. A second track, recorded by a second head, contains the chrominance signal as well as FM audio. The chrominance signal is a compressed time division multiplexed (CTDM) signal in which the analog R-Y and B-Y components are time compressed and then multiplexed into a single signal. Delaying the B-Y signal approximately one-half of a field allows the R-Y information to be recorded first, followed by the B-Y. FM audio is placed at the beginning of the track, before the color difference information. Upon playback, timing differences between the three component video signals are removed, and the components are time base corrected separately. The component signals are output at specified signal levels and also fed to an encoder. The encoder combines the signals to provide a composite output that can be used if desired.

One additional twist in many of the newer VTR formats is the use of compression or bit-rate reduction. Two formats, Ampex DCT and Sony's Digital Betacam perform a discrete cosine transform (DCT) on the video signal, reducing the number of bits required to approximately half. The DCT process is lossless, however, subsequent quantization can result in data being discarded permanently. Professional formats based on the consumer DV format such as Sony's DVCAM, Panasonic's DVCPRO and JVC's Digital-S (which uses a dual DV codec), as well as Sony's Betacam SX perform lossy compression on the video signal. The compression ratios used are minimal, and the resultant pictures are quite good, due in part to the fact they are recorded as digital rather than analog signals.

Audio Signal Path

The audio circuitry in videotape machines is comparable to the circuitry in professional audio recorders. One notable exception is a notch or filter around 15.7 kHz, the video horizontal frequency. Depending on whether the unit is intended for broadcast, industrial or consumer use, this is handled differently. One of the simplest methods is to simply roll off the audio response somewhere above 12 kHz such that the response at 15.7 kHz is relatively low. Other systems employ tighter notches specifically at the 15.7 kHz frequency. Many high quality professional units employ significant measures to prevent the 15.7 kHz from affecting the audio circuitry.

Most professional VTRs have balanced inputs and outputs referenced to +4 dBm. Most machines have at least two audio channels, while some have as many as four longitudinal channels. In addition, many offer up to two channels of pulse code modulated (PCM) or FM audio recording, however, on many units, these channels cannot be independently edited.

Quickly reviewing the audio record process, audio enters the machine, and once set at the proper level, is mixed with the bias current. The bias current is an ultrasonic (60 kHz to 200 kHz) low-distortion sine wave. Once the bias and audio signals are mixed, the output is fed to an equalization circuit for preemphasis. The signal is then recorded on tape. Upon playback, a deemphasis circuit restores the original frequency characteristics. Amplifiers then provide the required output from the machine. Most units provide E–E circuitry for monitoring the input signal(s) when the unit is in the stop mode.

ANALOG VIDEOTAPE FORMATS

The first successful videotape format was the 2″ quadruplex. Since then more than 30 videotape formats have been introduced. Only a few have been commercially successful in the professional arena. New tape formats, as well as older formats that remain in widespread use are covered here. For additional specifics, readers are referred to service information provided by the manufacturers.

3/4″ U-matic VCRs

The U-matic format is based on 3/4″ tape contained within a hard plastic shell (cassette). Two shell sizes are available, a large shell (which contains approximately 63 minutes of videotape) and a smaller shell (which can hold up to 23 minutes). A small slot in the bottom of the shells serves as a guide for insertion into machine elevator assemblies. When correctly aligned, the cassette lands on a set of fixed alignment pins within the transport. Loading mechanisms pull the tape from the cassette and wrap it around the drum assembly.

Reel tables are typically driven by direct drive motors or clutch and gear assemblies driven by a single motor. In most of today's machines, pulse generators within the reel tables provide feedback to control circuits. Lamps and optical pickups determine the beginning or end of tape. A clear leader at one end and clear leader with an opaque stripe at the other end provide feedback to ensure that sufficient tape is on each reel for the loading process. When either clear leader is encountered, tape movement is reversed, thereby ensuring that the remaining tape is not accidentally pulled from the hub assembly.

Within the transport assembly, when viewed from the front and above, the supply reel is on the right. Tape leaves the cassette shell and moves through several heads and guides. These typically include a full track erase head and one moveable guide (slanted) used to pull the tape from the cassette during threading. The tape rounds the slanted guide and begins its descent down the helix. In most U-matics the helix is a shelf machined into the lower drum assembly. The tape wraps the head drum assembly 180°. On the back side of the drum assembly the tape leaves the slanted guide but continues its descent, passing over the audio and control track (CTL) head and then to the capstan/pinch roller. The tape rounds a guide after the capstan, reversing its direction and beginning its ascent to the height of the take-up reel table.

At this writing, Sony is the only manufacturer still making 3/4″ machines. Today's machines offer improved performance (SP mode) and sophisticated machine control. Optional internal TBCs are available with outboard controls. Serial (RS-232, -422) and parallel remote control of all machine functions is common.

1″ Type C

1″ Type-C is an industry workhorse that remains in widespread use. Despite its popularity, use of type-C is declining. The format solved a major problem of the 2″ Quads—the use of multiple heads (4) during the active picture area. Differential response from the four heads caused a problem referred to as *banding*. The type-C format provides a single head that records one television field per scan. During playback there are no head switches and therefore no banding. The type-C format uses frequency modulation to record the entire composite video signal on each track. In addition to helically scanned video tracks, an area of the tape is set aside for optional helically scanned sync tracks. Longitudinal tracks include a control track and three audio tracks. Many type-C machines provide time code generators and readers for use with Audio 3. Vertical Interval Time Code (VITC) is also available.

The type-C format is an open reel, helical scan format. Servos include the head drum, capstan, reel motors and the video head itself. The type-C machines were the first to employ tracking servos to position the video playback head. Known by several names including Ampex's Automatic Scan Tracking (AST), RCA's SuperTrack and Sony's Dynamic Tracking, these servos used the off-tape RF envelope to determine the location of the tape's video track. Envelope tracking servos simplified machine operation, and allowed for tape playback without the need for control track information. Because of this, tapes could be played back at variable speeds. Depending on head flexibility, playback speed could range from −1X to +3X normal playback speed.

Betacam and M-II

The BetacamSP and M-II formats are highly successful refinements of earlier formats. M-II from Panasonic was an outgrowth of the less than successful M format. BetacamSP is an improved version of the Betacam format. Both M-II and BetacamSP require the use of metal particle tape. Both formats offer several advantages including ease of use, high quality component analog recordings, up to four channels of audio (2 longitudinal and 2 FM), and quality multi-generation editing. Most of these decks have built in TBCs standard, and several offer variable speed dynamic tracking. Camcorders and dockable recorders are available for field acquisition. Because of their portability and convenience, these formats are used in a wide variety of professional, semi-professional and broadcast applications. Two cassette sizes are available, and each shell type uses a different hub location. Because of this, models that accommodate both cassette shells must be capable of moving the reel tables. Depending on make and model one or both of the reel tables are moved to accommodate the different hub locations.

S-VHS & Hi-8

Both the S-VHS and Hi-8 formats are color-under, helical-scan formats designed for the high end consumer markets that have been adopted for professional use. Both offer backward compatibility, in that they can play and in many instances record tapes in the earlier mode (VHS and 8mm). Both Hi8 and S-VHS offer features for the professional market, including full editing capabilities, time code and dynamic tracking. JVC recently introduced a new innovation to the S-VHS format, control track time code. Rather than tying up an audio track for time code, the information is placed in the control track, and tapes can be *post-striped.* Post-striping allows time code to be properly added to the original recording, after the record process has been completed. Tapes can be shot in the field using a standard camcorder without time code, and post-striped in a studio deck. TBCs are available with Y/C inputs and outputs. Hi8 machines along with several other formats have added S-video connectors to the available I/O connectors.

DIGITAL VCR FORMATS

Digital recording formats have been replacing analog formats for the last 10–15 years. Although analog machines are still manufactured, no new analog formats have been introduced for nearly 10 years. Digital video formats (both tape and disk) have even found their way into the consumer video markets. In 1995, announcements concerning digital video decks for home use were made by several companies, followed by the DVD in 1997. Recording digital video is quite similar to recording digital data. Many of the differences between today's digital recorders parallel the differences found in the analog formats, and boil down to bandwidth and signal type.

D-1

The first digital videotape machines appeared in the mid-80's and were designated D-1. These component digital recorders were manufactured by Ampex and Sony and use 19 mm oxide tape and three different cassette shell sizes. The L-size can record 76 minutes of CCIR-601 video. The M-size can record 34 minutes and the S-size only 11 minutes. Video components are digitized and recorded using 8 bits and without compression. D-1, despite its age, remains the top of the line digital recording standard and is used throughout the post-production industry for multi-generational editing.

D-2

This composite digital format uses 19 mm metal tape, with machines originally available from Sony and Ampex. Three shell sizes are available, containing maximums of 208, 94 and 32 minutes of tape. Sampling structure is based on 4Fsc and the decks have widespread usage in composite facilities. A series of data recorders are based on these machines. Like D-1, 8 bits are recorded without using compression.

D-3

In 1990, Panasonic introduced the D-3 composite recording format which combined the small size of M-II and Betacam with the advantages of digital recording. D-3 uses 1/2 in. metal particle tape and provides for two cassette sizes, the small cassettes hold a maximum of 125 minutes of tape and the large hold up to four hours of tape.

DCT

The Ampex DCT component recorder uses 19 mm tapes in specially designed shells. The shells come in three sizes with maximum tape lengths the same as for D-2. A parallel product, DST, is a data storage unit based on the same technology. The units are similar, with the most obvious difference being the I/O modules. A mild 2:1 DCT compression is used but is essentially lossless. Ampex claims that error correction is used rather than error concealment, making the deck robust enough for data applications, and allowing perfect duplicates to be made from a master tape.

D-5

Panasonic's D-5 is a component recorder based on the D-3 transport and is backward compatible with D-3. An 18 MHz sample rate is included with the deck for making recordings of 16:9 aspect ratio material. Panasonic has made an adapter available that allows D-5 recorders to record the 1.2 Gb/s HDTV signal. D-5 decks record digital video at full data rates without compression, however, 4:1 compression is used for HDTV recordings.

Digital Betacam

Digital Betacam, from Sony, is a component format that uses 1/2 in. tapes along with a 2:1 DCT compression. These decks are capable of 10 bit component recordings, and some models are backward compatible with analog Betacam SP.

Betacam SX

Betacam SX provides 8 bit recordings on 1/2 in. tape, using approximately 10:1 compression. The MPEG 4:2:2 Studio Profile is used, as it provides for improved editability. Some of the Betacam SX decks include hard disk drives which allow a single machine to be used as an editor. The data rate used for Betacam SX recording is approximately 18 Mb/s.

Professional Derivatives of the Consumer DV Format

The consumer DV specification defines a recording and compression scheme based on 6.35 mm (1/4 in.) metal evaporated tape, the DV codec and a recording track pitch of 10 μm. No linear tracks are supported, as everything (video audio and data) is recorded digitally by the rotating heads in the scanner assembly. The signal structure used in the DV format is 4:1:1, 8 bit recording, and when combined with a compression ratio of 5:1 and the necessary data and error correction, produces a data rate of 25 Mb/s. Three professional recording formats have been derived from the DV Standard, Panasonic's DVCPRO, Sony's DVCAM and JVC's Digital-S.

Panasonic's DVCPRO uses 1/4 in. tapes and a 5:1 compression scheme. Transports are comparable in size to the half-height 5.25″ disk drive form factor. Consumer DV tapes can be played in the DVCPRO transports as can tapes produced on the Sony DVCAM format. DVCPRO requires metal particle tape because of two longitudinal tracks, a cue track and a control track, both designed to improve editing performance. DVCPRO recordings use a track pitch of 18 μm. A 4:2:2 version of DVCPRO, DVCPRO 50 is now available. A 100 Mb/s version of the format has been announced which will be capable of recording compressed HD signals.

DVCAM from Sony uses 1/4 in. metal evaporated tapes and a track pitch of 15 μm. No longitudinal tracks are recorded, however, a small IC within the shell is used to record cue points and other information to simplify the editing process. At present, DVCAM decks can playback DV consumer tapes, but not DVCPRO tapes.

Digital-S, from JVC uses 1/2 in. tape, and a dual DV codec to produce 4:2:2 recordings on an S-VHS compatible transport. Digital-S uses a compression ratio of 3.3:1 and some models can playback analog S-VHS tapes. Digital-S recorders at 50 Mb/s and a 100 Mb/s version has been announced.

HD Recorders

As professional video moves into the realm of High Definition (HD), recorders will be needed for these signals. Several recorders have already been implemented for HD. UniHi is an analog format that uses 1/2 in. metal tape to record 1125/60 signals. The UniHi luminance channel is limited to 20 MHz and the chroma difference channels are 7 MHz each. UniHi provides 4 channels of 16 bit digital audio. Sony's HDD-1000 is an open-reel 1″ digital machine that records a full bandwidth (30 MHz) HDTV signal. A tape speed of just over 30 ips is required to make uncompressed recordings of the 1.2 Gb/s signal. D-6, introduced by BTS/Toshiba also records full bit rate HDTV and uses the 19 mm D-1 cassette.

Compressed HD recordings are possible with a modified version of the D-5 format from Panasonic which uses 4:1 compression to record the HD signal at 300 Mb/s on 1/2 in. tape. HDCAM, from Sony is a HD camcorder and a companion studio editor that records what amounts to a 3:1:1 HD recording on 1/2 in. metal particle tape. Other HD recorders are available from JVC at both a consumer and industrial level.

VTR ACCESSORIES & OPTIONS

VTRs are used for storage and editing. However to be truly functional, the basic VTR requires some additional features. Although many models are available with all the bells and whistles, not all are. What follows is a list of some of the things that may not be included in all decks, but may be helpful—or necessary—in some applications.

Dynamic Tracking

Dynamic tracking was developed during the late 70's and first implemented on the type-C format. Since then it has found its way into the majority of helical scan formats. The concept is simple, however, the implementation is quite complex. Starting with the simplest example, if a non-standard tape is played back on a machine that is properly aligned, the playback tracking must be adjusted to center the playback head on the recorded track. With dynamic tracking, rather than adjusting the tracking control, the head is dithered (moved up and down) slightly, increasing and decreasing the RF envelope as the head moves toward and away from the center of the recorded track. By offsetting the head, based on the location of the increased RF envelope, the process can be used to consistently locate and center the head on the recorded track. A servo loop, using tip position and RF envelope amplitude, coupled with a dither signal (~450 Hz) can be used to locate the recorded tracks and follow them precisely.

To accomplish dynamic tracking, the head is mounted on a bimorphic strip (see Figure 5.7-4). The strip deflects when a sufficient voltage is applied (up to 400 V). Along one side of the strip, a sense strip is mounted. The sense strip provides feedback information to the servos. Use of the sense strip is less common today, as sufficient information can be obtained directly from the RF envelope. Bimorphic strips composed of a single element cause zenith errors. To reduce these errors, the bimorph is split near the end and the polarity is reversed (as is the sense strip, if used), causing the bimorphic distortion to become somewhat "S" shaped (see Figure 5.7-5).

Dynamic tracking can also be used to produce quality pictures at speeds other than normal play speed. For a standard head, recovering the recorded information becomes more difficult as the tape speed changes, due to the apparent change in track pitch. Dynamic tracking systems can handle these apparent changes in track pitch easily. Depending on flexibility, many dynamic tracking systems can playback stable, noise-free pictures at speeds ranging from −1X to +3X normal play speed. Some systems are capable of this range in a single continuous ramp, while others have discrete speeds stored in lookup tables. In addition to following the recorded track, the dynamic tracking servo must determine the appropriate times to jump or repeat tracks. In stop mode a single field is repeated constantly. At half normal speed each field must be repeated once. At 2X normal play speed, only every other field is played. At speeds other than normal play, the control track is not used nor is it required on most units.

Dynamic tracking adds to the work required from the TBC. Fields are reproduced from the tape as needed,

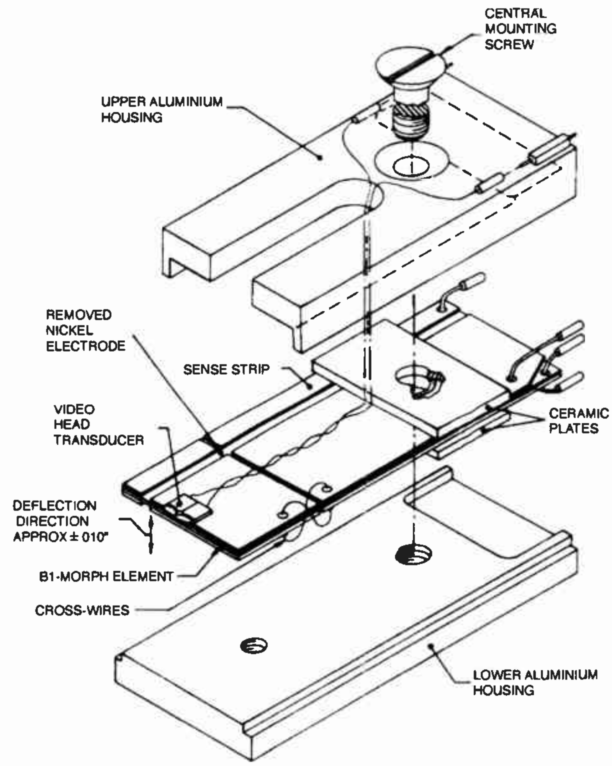


Figure 5.7-4. The Ampex AST dynamic tracking head used in 1" type-C machines. Similar units are common in many of today's professional and semi-professional helical scan machines.

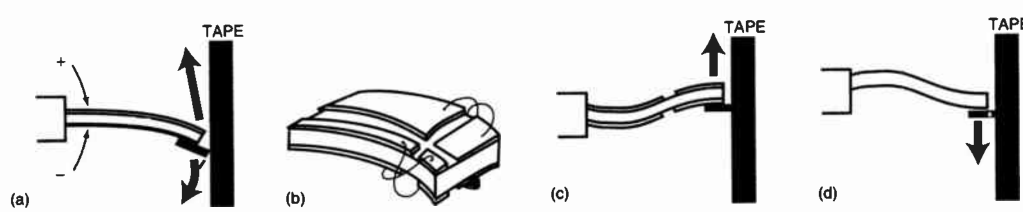


Figure 5.7-5. Video head deflection of dynamic tracking systems is accomplished by applying a voltage to the bimorphic strip the head is mounted on. If a single strip is used, (a) the deflection causes a zenith error. By adding an additional bimorphic segment (plus the sense strip), (b) the head deflection becomes a more "S" shaped (c & d) resulting in reduced zenith errors.

however, they generally do not agree with the sequence required for output. For instance, when playing at half speed, even or odd fields are played back to back and must be combined with a sync signal that includes both even and odd fields. Because of this, for every other field, the lines that make up the picture must be displaced by a single line. As the picture is displaced up and down it results in a phenomenon known as *field hop*. At half speed field hop is not noticeable, however, at speeds within 5% of play speed, as well as others, it can become objectionable.

Time Base Correction

Time base correction is required to make all analog tape machines meet RS-170(A) specifications. For digital machines, this can be accomplished using buffers and error correction routines. Time base errors are caused by several factors including tape stretch, temperature induced expansion and contraction of the tape binder, and by minute distortions and ripples of the tape surface caused by head impacts. Off-tape horizontal sync pulses and color burst phase are typically used to determine the amount of error in the signal. Today's TBCs employ various amounts of memory, the most common being 16 horizontal lines for a time base corrector and an entire field or frame for frame synchronizers with time base correction capability. Vertical sync, typically advanced by an amount equal to half the buffer (8 lines) is fed from the TBC to the tape machine, to allow time to process the signal. Frame synchronizers, because of their increased memory, simply delay the signal to align the incoming video signal with the reference sync. A frame synchronizer will not do TBC and a TBC will not do frame synchronization, however, the circuitry involved is similar and many times both units are found in a single package. Many of these single unit packages will automatically switch to whatever mode is required, based on the input signal.

Today's digital TBC/frame synchronizers are simple, rather elegant, solutions to what used to be a very difficult problem. In these units, the video information is converted to digital by sampling at a frequency normally three or four times the subcarrier frequency (3.58 MHz). Once the picture information is in a digital form, it is clocked into memory based on the incoming (off-tape) horizontal sync pulses. Later the information is clocked out of memory using the reference sync pulses. In this manner, TBC simply happens. The amount of information stored in memory shrinks and expands as necessary. Ideally, memory remains 50% full at all times. An error occurs if the read and write circuits attempt to access the same location simultaneously. This can happen both when the memory is empty and when it is full. Special effects such as slow motion can tax the system, and machines that use these effects generally require TBCs with larger memory buffers.

Field hop was discussed earlier and was addressed in the *Zeus* series of TBCs from Ampex. The problem of field hop occurs because the interlaced NTSC system requires two fields vertically displaced by a single horizontal line. If, during slow-motion playback, an even field was output on an odd field, the picture would be shifted vertically one line. Later when the field sequence returned to normal, the picture shifts in the opposite direction. If these shifts occur within a few seconds of each other, the picture appears to hop up and down. The Zeus system addresses this by displacing all fields by one-half of a line during dynamic tracking sequences. This way the picture can be held stable vertically.

Time Code

Several versions of time code exist, but the one most commonly associated with videotape machines is SMPTE time code (described in ANSI standard V98-12). Originally longitudinal time code (LTC) was recorded on an address track if available, or an audio track. The problem with all longitudinal recordings is the increased difficulty reading them as the tape speed drops. Since the primary use of time code is for editing, it is very important to accurately locate cue points. Precisely parking the tape for preroll becomes very difficult, as the nearer the machine is to the cue point, the slower the tape moves. One method is to simply count control track pulses if time code becomes unreadable. Unfortunately, control track is also a longitudinal recording. Vertical interval time code (VITC) addresses this problem. VITC is recorded on one or two horizontal lines in the vertical interval and is readable as the tape slows, because the video head remains in motion. Machines with dynamic tracking are helpful when reading VITC at slow speeds, but are not required. The proper line in the vertical interval is decoded and provides time code information at slow tape speeds.

When both VITC and LTC are recorded on a tape, the tape can be searched and cued quickly and efficiently. Machine detection circuitry normally uses VITC at speeds below ½X play speed. At speeds above 2X play speed, LTC is used. LTC can be read at speeds well over 25X play speed. Between ½X and 2X play speed, either VITC or LTC can be used, but most machines default to LTC. It is important that both LTC and VITC agree. Although it should be impossible to record two different time codes, it seems to be a common problem. A breakdown of the time code bit structure is shown in Figure 5.7-6.

MAINTENANCE

Today's tape machines are highly reliable units. The biggest problems are mechanical, and are generally caused by the accumulation of dirt on the mechanical transport assemblies. Units should be cleaned regularly with lint-free cloths and non-residue producing solvents, taking care not to bend or in any way damage or distort the various guides. Pinch rollers should be inspected as they will distort with age and use. Video head assemblies should be cleaned carefully as the tips are fragile. A good method is to hold the cloth against the drum and rotate the drum slowly in its normal

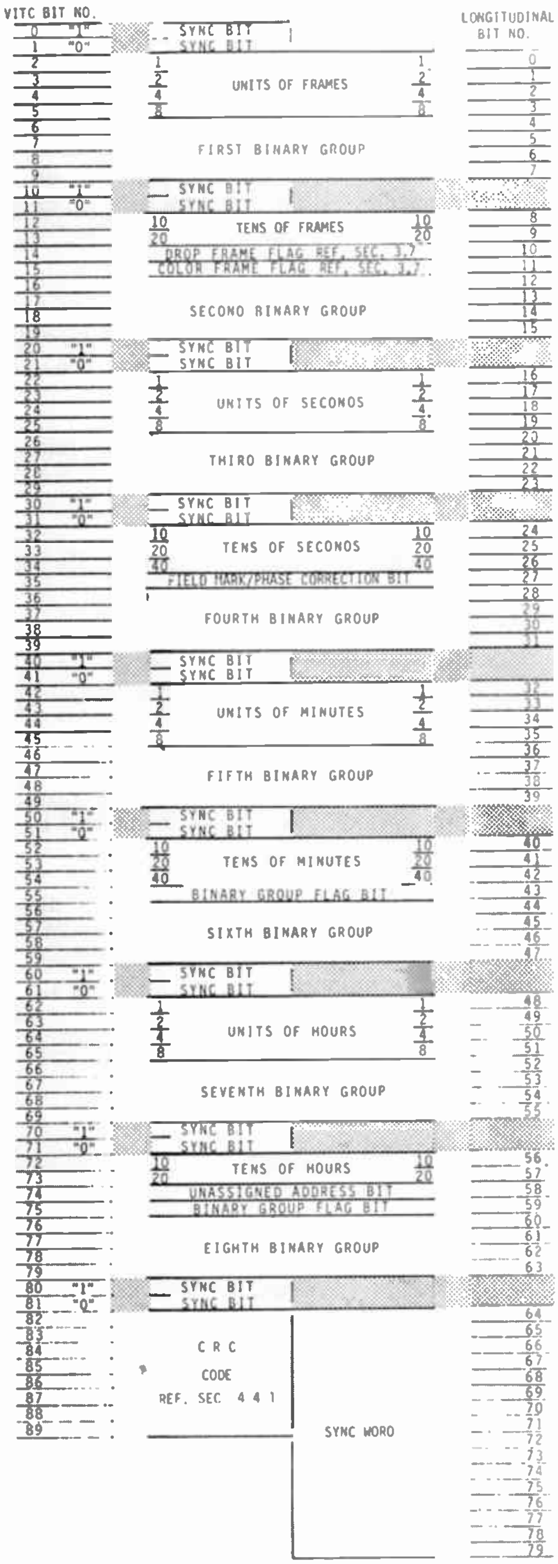


Figure 5.7-6. Bit structure of SMPTE time code, 80 bits are used for longitudinal time code, 90 bits are used for VITC. Although they are part of the standard, the user bits have received only limited use to date. (Source: *NAB Engineering Handbook, Eighth Edition, p. 921*)

direction of travel, leaving the cloth stationary. Mechanical assemblies in the tape path should be carefully inspected for wear as well as scratches or nicks that could damage the tape. Any parts showing excessive wear should be replaced. Occasional checks should be made to verify the unit is on standard. This is accomplished using an alignment tape. Before any attempt is made to service a tape machine, consult the manufacturer for recommended procedures, as significant damage can result if done improperly.

When aligning most types of recording equipment, the procedure is to playback a recording that is known to be on standard. In the case of VTRs, these recordings are in the form of alignment tapes. Tapes are typically produced on *golden machines* that are verified to be on standard through various methods which include precision mechanical devices. Once recorded, alignment tapes are considered usable for approximately 100 passes, at which time they should be replaced. Generally, by that time, the tapes have been stretched and/or damaged to an extent that the recordings are no longer standard. Various sections of the tape are used to verify proper operation of different portions of the videotape machine. Although, it is possible to copy an alignment tape, the new recording is only as good as the machine that was used to record it.

Normally, each tape contains an interchange section, a video section and an audio section. Cue tones, time code and other signals used within the format are also recorded on the tape. This allows the various mechanical and electronic sections to be properly aligned. The interchange section is used to verify proper mechanical alignment of the transport and servos. Video and audio sections provide test signals for setting the proper playback levels and equalization. Once a machine is setup for proper playback of the alignment tape, a recording is made and checked to see if it plays back properly. Any problems found at this stage are due to errors in the record section, which can then be adjusted so that recordings made on the machine will play back properly on any unit that is on standard.

DISK RECORDERS

Analog recorders using disk technology were first introduced in the '70s. The units were physically large and bulky, while the available recording time was very small. Applications included slow-motion playback for sports. Today, a variety of units are available. Most are based on magnetic recording technology, while some use optical methods or a combination of the two. Some disk recorders using optical technology are rewriteable, however, most optical methods are of the write-once variety. Although some older magnetic units provided removable media, the media within nearly all of today's magnetic drives designed for video use is non-removable. Most optical based drive systems are removable. The primary advantage of disk technology is the ability to randomly access individual frames—almost instantly. Videotape requires the tape to be shuttled—in a linear fashion—to the desired frame or clip. Depending on the clip's location, and the speed of the transport, the time required can be considerable. Video disk recorders are widely used in editing environments and for applications such as news and master control that make repeated use of short video and audio clips.

Most of today's video disk recorders are based on hard drive technology developed for the computer industry. To obtain the necessary storage capacity and throughput, many individual drives are combined and function as a single unit. Most systems make some use of compression, which reduces both bandwidth and storage requirements. Although hard drives require little maintenance and have MTBFs quoted over 50,000 hours, their failure can be disastrous. Therefore, many systems provide some type of redundancy. A common redundancy scheme is RAID (redundant array of inexpensive drives). Various RAID levels are defined, and include:

- RAID 0—a combination of multiple drives that has no redundancy.
- RAID 1—disk mirroring, where two or more drives are configured such that redundant drives are mirrored. All data is stored on each of the drives.
- RAID 2—uses additional error correction systems and is not widely implemented.
- RAID 3—uses one additional drive for parity information.
- RAID 4—is similar to RAID 3, but redistributes the information to better accommodate multiple simultaneous reads.
- RAID 5—uses one additional drive, but distributes parity information across all drives.

RAID 3 and RAID 5 are the levels most used for video applications. In both, the loss of a single drive will not result in any data loss, however, the loss of a second drive before the first drive is rebuilt results in a complete loss of any data not yet rebuilt.

As disk recorders for video mature, many are finding their way into server applications. In these applications, video is stored on drives and accessed using techniques and systems comparable to computer file servers. (See Chapter 5.8 *Video Server Storage Systems* for more information on servers and disk based systems. Other disk based recorders are being used as direct replacements for VTRs. In these applications, the disk recorders emulate many, if not all, of the functions commonly found in tape machines. A few of these units are bridging the computer and video worlds. These units can act like VTRs as well as being addressable as network drives. This allows users to move information easily between computer-based systems to traditional video systems. Not all video disk recorders are designed to record audio, some units are video only. When considering video disk systems, items such as the use of compression, image quality, storage capacity and interface requirements need to be carefully reviewed.

CONCLUSION

Videotape recording has come a long way since its introduction. Decks have gained considerable functionality and provide many of the essential tools for broadcast and professional applications. Although the days of the VTR are far from numbered, disk recorders are providing many of the VTRs capabilities alongside a new generation of computer based non-linear production technology. Tape will likely remain a viable acquisition technology for years to come, as disk acquisition is still in its infancy. Analog formats are also likely to remain in use for many years (consider how long 2″ survived), however, the majority of new recordings will be made using digital decks or onto disk recorders. As VTRs move from the center of the editing process, to devices used only at the beginning and end of the process, the wear and tear will be reduced, extending the lifetime of these machines. Later, as server technology improves, VTR use is likely to decline further. VTRs will be needed for many years for archive playback. Regardless, of how antiquated videotape becomes, tape recording technology will be necessary. Servers and computer systems require regular backup, and tape remains the most cost effective large scale storage available.

Although some have said they have purchased their last VTR, they remain the most cost effective and reliable storage method for most video programming longer than a few minutes. However, as the size of the transports is reduced, the amount of maintenance possible within the average facility will also be reduced. Electronic components are getting smaller, as are the parts counts on the circuit boards. More and more, small format transports are likely to become disposable, or not user serviceable, reducing the cost effectiveness of in-house repair. Tape is likely to see a resurgence over the next few years for HD recording, but that may be short lived. With minimal compression, today's computer disk drives are capable of handing video efficiently. If Moore's law holds, drive technology is likely to be capable of handling HD recording by the year 2000 or shortly thereafter.

KEY TERMS

Color-under recording. A recording process that adds an AM chrominance subcarrier centered near 650 kHz to an FM luminance carrier for recording on tape. This process is used for some professional formats and most consumer formats.

Control track (CTL). One of several longitudinal tracks laid down during the recording process. Information from the CTL is used to position the tape for proper playback. In many ways, the control track is the electronic equivalent of sprocket holes in film.

Dithering. The process of moving the video head perpendicular to the recorded track. An oscillator provides a 450 Hz signal to the head assembly to move it in such a way as to locate the recorded track precisely.

Under servo control, the head can deflect a precise distance and remain centered on the track even at speeds other than normal play speed.

Drum assembly. The drum assembly in helical scan machines typically consists of an upper and lower drum. The upper drum contains the head and preamp assemblies. The lower drum consists of the reference edge guide (helix), rotary transformers and the servo motor used to rotate and position the video heads. The lower drum also contains a pulse and/or frequency generator that provides rotational velocity and position information to the drum servo.

Helical scan. A recording method in which the tape is wrapped around a rotating drum in the shape of a helix. Various methods exist which vary drum diameter, number of recording heads, wrap angle and direction of tape travel (either the same or opposite direction as drum rotation).

LTC. Longitudinal Time Code

Quads. Video tape machines that used 2″ tape and the transverse scan recording method pioneered by the Ampex Mark IV in 1956. The format was referred to as *quadruplex* because four heads, mounted 90° apart on a drum assembly, were used to record the video signal on tape.

Time code. A system of numbering the individual frames of a video recording. Valid time code numbers run from zero to 23:59:59:29, a 24-hour clock. Time code can be generated as an audio signal and recorded longitudinally (LTC) on audio tracks, or can be inserted into video on a single horizontal line of the vertical interval (VITC). Many small format recorders insert two identical lines of time code due to the possibility of a dropout making a single line unusable.

Transverse scan. A recording method in which the recorded track is perpendicular to the direction of tape travel.

VITC. Vertical Interval Time Code

BIBLIOGRAPHY:

Ampex Training Department, *General Information, Volume 1,* Ampex Corporation, 1983.

Benson, K. B., and Whitaker, Jerry., *Television Engineering Handbook, Revised Edition,* McGraw-Hill, 1991.

Benson, K. B., and Whitaker, Jerry., *Television and Audio Handbook for Technicians and Engineers,* McGraw-Hill, 1990.

Epstein, Steve., Videotape Storage Systems, *The Electronic Handbook,* 1996, CRC Press, Boca Raton, FL.

Ginsburg, Charles P., *The Birth of Videotape Recording,* Ampex 1981, Reprinted from notes of a paper delivered by Ginsburg to the SMPTE on Oct. 5, 1957.

Hammer, Peter., The Birth of The VTR, *Broadcast Engineering,* June 1986.

Roizen, Joseph., Magnetic Video Recording Techniques, *Ampex Training Manual—General Information, Volume 1.*

Mee, C. Dennis and Daniel, Eric D., *Magnetic Recording Handbook,* McGraw-Hill, New York, 1990.

NAB Engineering Handbook, Sixth (1975), Seventh (1985) and Eighth (1992) editions.

Sanders, Mark., Technology Report: AST, *Video Systems,* April 1980.

Additional information on the topic of videotape machines is available from the following sources:

Broadcast Engineering, a monthly periodical dealing with television technology. The magazine, published in Overland Park, KS, is free to qualified subscribers.

The Society of Motion Picture and Television Engineers, which publishes a monthly *Journal,* and holds an annual technical conference in February and a convention in the Fall. The SMPTE is headquartered in White Plains, NY.

5.8
VIDEO SERVER STORAGE SYSTEMS

JERRY WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

Video servers, with their unique set of features and functions, have reshaped the way program segments are stored and played to air. These systems have emerged from the realm of limited use, special purpose devices to mainstream video production. Among the many attributes of video servers are:

- They permit material from a single storage source to be used simultaneously by multiple users
- Provide a migration path to the all-digital facility that is not necessarily format limited
- Result in a reduction in lost or misplaced materials
- Reduction in the size and space requirements relative to a tape environment
- Complete computer-control capabilities
- Near instant access and playback of video segments

The end result of these attributes is an environment where multiple applications and/or services can be generated from a single system. This reduces the amount of playback equipment required, reduces tape consumption and generally permits more efficient use of human resources.

Architecture

The basic video server architecture consists of three elements:[1]

- A multiple hard disk drive system capable of fast and simultaneous data access, with sufficient capacity and redundancy for the contemplated application. A disk array controller manages data distribution and communications between all drives
- Fast data communication interfaces between disk drives and networks. Several approaches can be used to perform fast data transfer such as very fast CPUs, multiple CPUs and buses, and routing switchers. Interfaces may also include data compression encoding and decoding
- An operating system capable of handling multiple digital audio and/or video data streams in any combination of record and playback modes, while ensuring correct file management and easy access

The design and performance of the server involves trade-offs between the quality of compressed signals, storage capacity, data speed, play time, number of channels, access speed and reliability. Table 5.8-1 shows the storage space requirements for different audio and video signals.

All video servers have in common a large storage capacity and multiple channel capability; differences involve the basic architecture and performance targets, because different systems may be designed to meet specific requirements. The most common application groups for broadcasting include the following:

- **Transmission.** High compression ratios can be used for server-to-air applications, typically commercial and short-length program replay. The resulting bit rates and number of channels are usually low, therefore, with relatively low bandwidth requirements
- **Data cache.** A caching system is basically a temporary random access disk buffer. It is commonly used in conjunction with tape library systems in on-air applications. Caches are well suited to commercial on-air insertion applications where elements are repeated several times a day and last moment changes occur frequently. Bandwidth and channel requirements are usually not critical
- **News.** For most news materials, moderate overall quality levels are acceptable. Video compression is used to reduce the data file size and increase the transfer rate. Multiple access to multiple segments must be possible and a moderate bandwidth is necessary. Guaranteed availability of output ports might be required for direct on-air programming of news materials
- **Production.** Wide bandwidth must be provided to meet simultaneous transfer demands for large uncompressed video files with real-time and random access capabilities. In a production facility, a server stores all compressed or uncompressed audio and

Table 5.8-1
Storage space requirements for audio and video data signals.

Media Signals	Specifications	Data Rate
Voice-grade audio	1 ch; 8 bit @ 8 kHz	64 kbps
MPEG audio Layer 2	1 ch; 16 bit @ 48 kHz	128 kbps
MPEG audio Layer 3	1 ch; 16 bit @ 48 kHz	64 kbps
AC-3	5.1 ch; 16 bit @ 48 kHz	384 kbps
CD	2 ch; 16 bit @ 44.1 kHz	1.4 Mbps
AES/EBU	2 ch; 24 bit @ 48 kHz	3.07 Mbps
MPEG-1 (video)	352 × 288, 30 fps, 8 bit	1.5 Mbps
MPEG-2 (MP@ML)	720 × 576, 30 fps, 8 bit	15 Mbps, max.
MPEG-2 (4:2:2 P@ML)	720 × 608, 30 fps, 8 bit	50 Mbps, max.
ITU-R Rec. 601	720 × 480, 30 fps, 8 bit	216 Mbps
HDTV	1920 × 1080, 30 fps, 8 bit	995 Mbps

Table 5.8-2
Production and broadcast server bandwidth requirements for 4:2:2 video signals.

Production Applications	Bandwidth Mbps	Sample Resolution	Compression Ratio
High-end post-production	270	10	1
Typical post-production	90	10	2.3:1
Low-end post-production	25–50	8	6.6:1–3.3:1
News (compressed data)	18–25	8	9:1–6.6:1
HDTV broadcast	20	8	10:1–50:1
Good-quality SDTV broadcast	8	8	20:1
Medium-quality SDTV broadcast	3	8	55:1
Low-quality SDTV broadcast	1.5	8	110:1

video data files for use in post-production and distribution. The server is central to the production operation and, thus, determines the overall performance of the facility. Table 5.8-2 lists production server bandwidth requirements for 4:2:2 video signals
- **Video on demand (VOD).** This server application must deliver a large number of channels, each of relatively low video quality, such as MPEG-1 (1.5 Mbps). A high overall bandwidth may be required to satisfy all demands. Short access times are necessary for VOD, whereas long access times permit *near-VOD* (NVOD) only.

Video servers have redefined many of the common applications of videotape recorders (VTR). The VTR is being displaced in direct-to-air uses, complex post-production and desktop editing. As new VTR formats evolve, they will increasingly be aimed at acquisition and long-form storage applications. Table 5.8-3 lists the digital tape formats in common usage.

Table 5.8-3
Digital video tape formats.[3]

Format Designation	Format Description
D1	A format for digital video tape recording conforming to the ITU-R Rec. 601 (4:2:2) standard using 8 bit sampling. The tape is 19 mm wide and allows up to 94 minutes to be recorded on a cassette.
D2	The VTR standard for digital composite (encoded) NTSC or PAL signals. D2 uses 19 mm tape and records up to 208 minutes on a single cassette. Neither cassettes nor recording formats are compatible with D1. D2 has often been used as a direct replacement for 1 in. analog (Type C) VTRs.
D3	A VTR standard using 1/2 in. tape cassettes for recording digitized composite (encoded) NTSC or PAL signals sampled at 8 bits. Cassettes are available ranging from 50–245 minutes.
D4	There is no D4 designation; most DVTR formats were developed in Japan where "4" is regarded as an unlucky number.
D5	A VTR formal using the same cassette as D3 but recording component signals sampled to ITU-R Rec. 601 at 10 bit resolution. With internal decoding, D5 VTRs can play back D3 tapes and provide component video outputs.
D6	A helical-scan digital tape format that uses a l9 mm cassette to record uncompressed HDTV material at 1.88 Gbps. D6 accepts both the European 1250/50 interlaced format and the SMPTE 260M version of the 1125/60 interlaced format, which uses 1035 active lines.
D7	The designation assigned to the DVCRPO recording format.

Figure 5.8-1. Simple application of a video server in a one in, many out configuration.

Among the technical frontiers for video servers is the ability to transfer files at faster than real-time. Such features place considerable demands on the bandwidth and throughput of the system; however, they provide many operational benefits, such as accelerated nonlinear editing and reduced program load-up times.

Video Server Design

A number of operational scenarios can be implemented with a video server system. The simplest involves multiple streams of one or two video inputs, as illustrated in Figure 5.8-1. All the material is stored within the server and plays out to multiple channels as required.

In Figure 5.8-2, the basic system is expanded to allow for some material, particularly long form programs, to be held in an external storage area or device (most likely on tape) and loaded into the server on an as-needed basis to coexist with the short-form (typically commercial) library that is stored in the server. The server then manages the entire play-out process to one or more outputs.

In Figure 5.8-3, a large-scale system is depicted typical of a newsroom environment. Feeds are received into the system through a triage station permitting rough editing of material (*keep or discard*). The selected material is then stored in a feed server, which is accessible for viewing or by editors for story composition. Completed program elements are forwarded to the on-air server, which stores them and plays-out as needed. Also shown in the Figure is a *clip server*, a low resolution companion server that delivers video to multiple desktop viewing and editing systems.

Video servers are often constructed from off-the-shelf computer components by integrating the basic hardware and software. This approach promotes a migration path for future growth in several ways, including:

- Utilization of standardized components
- Enhanced ability to expand or alter the system
- Ability to upgrade system core capabilities with new software releases

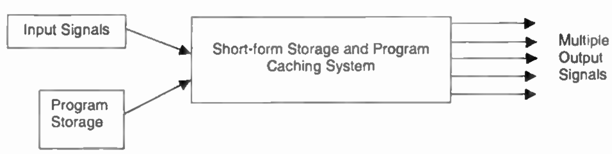


Figure 5.8-2. A server system expanded to utilize long-form program storage.

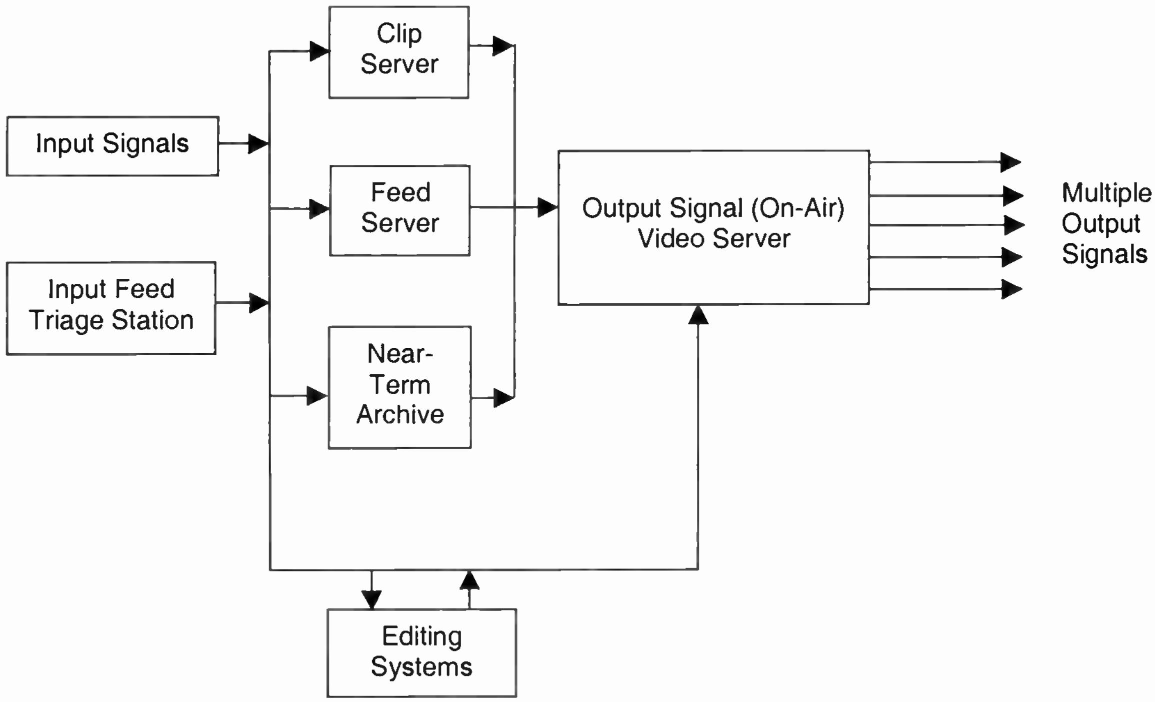


Figure 5.8-3. A server system designed for editing-intensive applications involving a variety of program inputs and distribution requirements.

By utilizing standard components, the research and development cost are spread across a large market, providing faster and more cost-effective advancement. For example, all disk drives adhere to a similar footprint, respond to a similar set of commands, mount in a similar manner and so forth. This makes it easy to take advantage of improvements in specific technologies.

Archiving Considerations

As a conventional tape library grows, a two-sided challenge unfolds. First, operators must identify and archive the contents of the tape. Next, they must be able to find the tape. Servers inherently must maintain a database of the materials within their contents and manage the recall of that material on demand. In addition, because the server brings basic data storage concepts into play, more use is being made of asset management software that permits identification and location of all materials and their contents, regardless of the storage medium.

Tape, however, remains an exceptionally robust and inexpensive medium for field acquisition. It also is economical for many long-form materials and for archiving. Tape and servers can work together to provide a performance and economic model that offers maximum benefit to the user. As illustrated in Figure 5.8-4, raw footage is acquired on tape. The tape material is loaded into the editors for composition and the finished programs are forwarded to a server, where it is held for play-out.

Despite the many benefits of server systems, there are a few drawbacks. Catastrophic failure is first among them. Tape, as a medium, is not susceptible to catastrophic failure (generally speaking). Disk drives can—and do—crash. However, advancements in drive technology have greatly increased the mean time between failure for disk drives. *Redundant array of independent disks (RAID)* schemes provides built-in backup, particularly when linked with hot spare drives and hot-swappable components such as power supplies. With these additions, the risk of total failure of a server system is quite low.

VIDEO SERVER STORAGE SYSTEMS

Because of the fundamental impact that storage technology has on any video server, it is important to understand the various RAID configurations commonly in use. The typical solution to providing access to many gigabytes of data to users fast and reliably has been to assemble a number of drives together in a *gang* or array of disks. Simple RAID subsystems are basically a group of up to five or six disk drives assembled in a cabinet that are all connected to a single controller board. The RAID controller orchestrates read and write activities in the same way a controller for a single disk drive does, and treats the array as if it were—in fact—a single or *virtual* drive. RAID management software, which resides in the host system, provides the means to control the data being stored on the RAID subsystem.

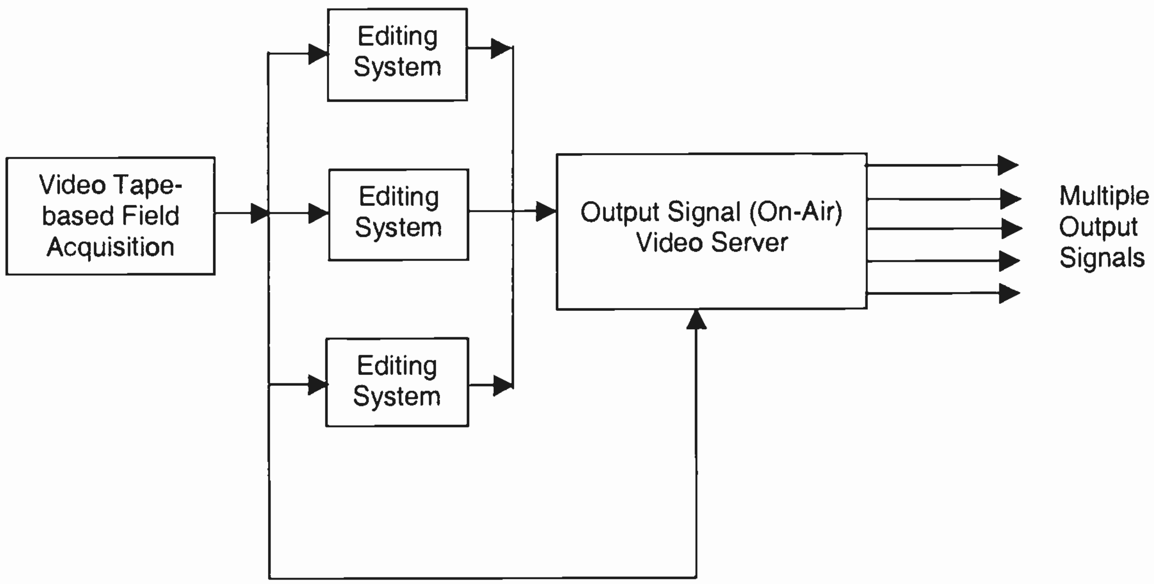


Figure 5.8-4. A video tape-to-server environment with editing capabilities.

Despite its multi-drive configuration, the individual disk drives of a RAID subsystem remain hidden from the user; the subsystem itself is the virtual drive, though it can be as large as 1,000 Gbytes or more. The phantom virtual drive is created at a lower level within the host operating system through the RAID management software. Not only does the software set up the system to address the RAID unit as if it were a single drive, it allows the subsystem to be configured in ways that best suit the general needs of the host system.

RAID subsystems can be optimized for performance, highest capacity, fault tolerance, or a combination of these attributes. Different *RAID levels* have been defined and standardized in accordance with these general optimization parameters. There are six such standardized RAID levels, called RAID 0, 1, 2, 3, 4 and 5. The use of a particular level depends upon the performance, redundancy, and other attributes required by the host system.

The RAID controller board is the hardware element that serves as the backbone for the array of disks; it not only relays the input/output (I/O) commands to specific drives in the array, but provides the physical link to each of the independent drives so they may easily be removed or replaced. The controller also serves to monitor the integrity of each drive in the array to anticipate the need to move data should it be placed in jeopardy by a faulty or failing disk drive (a feature known as *fault tolerance*).

Basic Drive Technology

Small computer system interface (SCSI) hard disk drives are the foundation of any RAID system. The drives use various numbers of magnetic coated disks that rotate at 5,400, 7,200, or 10,000 RPM.[2] Each disk has two recording surfaces. The disks are logically divided into concentric circles (*tracks*). A set of tracks at a given position on the disks is known as a *cylinder.* The number of cylinders is the same as the number of tracks across the disks. Tracks are then divided into varying numbers of *sectors.* The number of sectors varies according to the position of the track located on the disk. This technology is known as *zone bit recording* (ZBR).

Data is written to one track of a given surface at a time. If more space is needed, a head switch takes place and the data is then written to the next surface. Head switches continue down the cylinder until the last track in the cylinder is filled. If additional space is needed, the heads are stepped to a new cylinder and the process of head switching and track stepping continues until the file is completed. Using ZBR, the outside tracks contain more sectors than the inside tracks because there is more physical space on the outside tracks. The transfer rate is greater from the outside as well because a greater amount of data is available from a single rotation of the disks. The first partition created uses the outside tracks and moves to inside tracks as the drive or drives are fully partitioned.

SCSI is a general purpose interface. In its basic configuration, a maximum of eight devices can be connected to a single bus. SCSI specifies a cabling standard, a protocol for sending and receiving commands and the format for those commands. SCSI is intended as a device-independent interface so the host computer requires no details about the peripherals that it controls.

As the SCSI Standard has evolved, various levels have been produced, the most common being:

- Standard SCSI, with a 5 MBps transfer rate
- Fast SCSI, with a 10 MBps transfer rate

- Ultra SCSI, with a 20 MBps transfer rate
- Ultra SCSI-2, with a 40 MBps transfer rate

For each of these schemes, there is the 8 bit normal or so-called narrow bus (1 byte per transfer) or the 16 bit wide bus (2 bytes per transfer). Therefore, for example, Ultra SCSI-2 is designed to transfer data at a maximum rate of 80 MBps. Actual continuous rates achieved from the disk drive will be considerably less, however.

RAID Levels

The RAID 0 through 5 Standards offer users a host of configuration options.[3] These options permit the arrays to be tailored to their application environments, for broadcast purposes a video server. Each of the various configurations focus on maximizing the abilities of an array in one or more of the following areas:

- Capacity
- Data availability
- Performance
- Fault tolerance

RAID Level 0

An array configured to RAID Level 0 is an array optimized for performance, but at the expense of fault tolerance or data integrity.[3] RAID Level 0 is achieved through a method known as *striping*. The collection of drives (*virtual drive*) in a RAID Level 0 array has data laid down in such a way that it is organized in stripes across the multiple drives. A typical array can contain any number of stripes, usually in multiples of the number of drives present in the array. As an example, a four-drive array configured with 12 stripes (four stripes of designated *space* per drive). Stripes 0, 1, 2 and 3 would be located on corresponding hard drives 0, 1, 2 and 3. Stripe 4, however, appears on a segment of drive 0 in a different location than Stripe 0; Stripes 5 through 7 appear accordingly on drives 1, 2, and 3. The remaining four stripes are allocated in the same even fashion across the same drives such that data would be organized in the manner depicted in Figure 5.8-5. Practically any number of stripes can be created on a given RAID subsystem for any number of drives. Two hundred stripes on two disk drives are just as feasible as 50 stripes across 50 hard drives. Most RAID subsystems, however, tend to have between three and 10 stripes.

The reason RAID 0 is a performance-enhancing configuration is that striping enables the array to access data from multiple drives at the same time. In other words, because the data is spread out across a number of drives in the array, it can be accessed faster because it is not bottled up on a single drive. This is especially beneficial for retrieving a very large file, because it can be spread out effectively across multiple drives and accessed as if it were the size of any of the fragments it is organized into on the data stripes. By any measure of comparison, video files qualify as very large files.

The downside to the RAID Level 0 configuration is that it sacrifices fault tolerance, raising the risk of data loss because no room is made available to store redundant information. If one of the drives in the RAID 0 fails for any reason, there is no way of retrieving the lost data, as can be done in other RAID implementations.

RAID Level 1

The RAID Level 1 configuration employs *disk mirroring* to ensure data reliability or a high degree of *fault tolerance*.[3] RAID 1 also enhances read performance, but the improved performance and fault tolerance come at the expense of available capacity in the drives used. In a RAID Level 1 scheme, the RAID

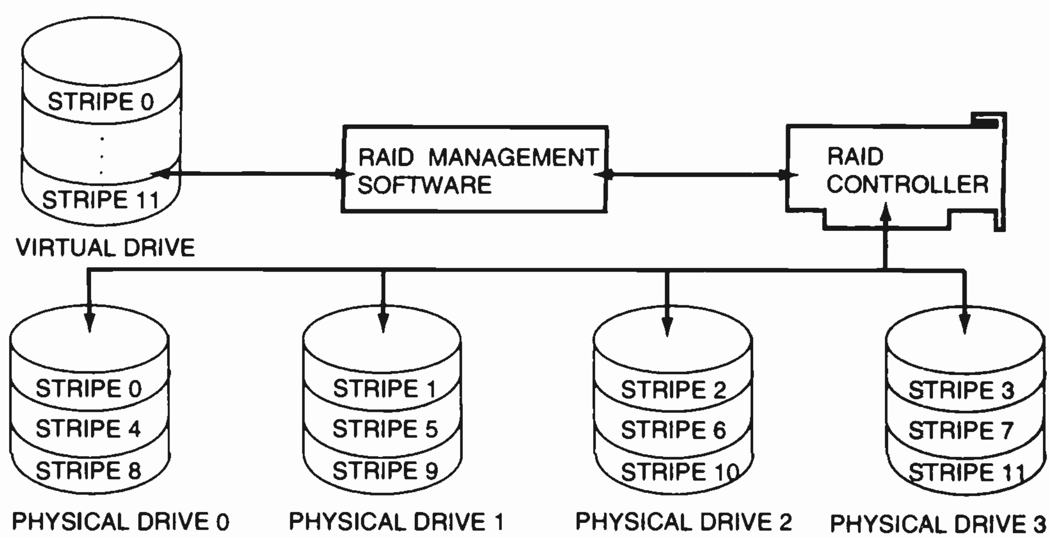


Figure 5.8-5. In a RAID Level 0 configuration, a virtual drive is comprised of several stripes of information. Each consecutive stripe is located on the next drive in the chain, evenly distributed over the number of drives in the array.[3]

management software instructs the subsystem's controller to store data redundantly across a number of the drives (*mirrored set*) in the array. In other words, the same data is copied and stored on different disks to ensure that, should a drive fail, the data is available somewhere else within the array. In fact, all but one of the drives in a mirrored set could fail and the data stored to the RAID 1 subsystem would remain intact. A RAID Level 1 configuration can consist of multiple mirrored sets, whereby each mirrored set can be a different capacity. Usually, the drives making up a mirrored set are of the same capacity. If drives within a mirrored set are of different capacities, the capacity of a mirrored set within the RAID 1 subsystem is limited to the capacity of the smallest-capacity drive in the set, hence the sacrifice of available capacity across multiple drives.

The read performance gain can be realized if the redundant data is distributed evenly on all of the drives of a mirrored set within the subsystem. The number of read requests and total wait state times both drop significantly; inversely proportional to the number of hard drives in the RAID, in fact. To illustrate, suppose three read requests are made to the RAID Level 1 subsystem (see Figure 5.8-6). The first request looks for data in the first block of the virtual drive; the second request goes to block 0 and the third seeks from block 2. The host-resident RAID management software can assign each read request to an individual drive. Each request is then sent to the various drives, and now—rather than having to handle the flow of each data stream one at a time—the controller can send three data streams almost simultaneously, which in turn reduces system overhead.

RAID Level 2

RAID Level 2, rarely used in video applications, is another means of ensuring that data is protected in the event drives in the subsystem incur problems or otherwise fail.[3] This level builds fault tolerance around *Hamming error correction code* (ECC), which is often used in modems and solid-state memory devices as a means of maintaining data integrity. ECC tabulates the numerical values of data stored on specific blocks in the virtual drive using a formula that yields a checksum. The checksum is then appended to the end of the data block for verification of data integrity when needed.

As data is read back from the drive, ECC tabulations are again computed, and specific data block checksums are read and compared against the most recent tabulations. If the numbers match, the data is intact; if there is a discrepancy, the lost data can be recalculated using the first or earlier checksum as a reference.

This form of ECC is actually different from the ECC technologies employed within the drives themselves. The topological formats for storing data in a RAID Level 2 array is somewhat limited, compared to the capabilities of other RAID implementations, which is why it is not commonly used in commercial applications.

RAID Level 3

This RAID level is essentially an adaptation of RAID Level 0 that sacrifices some capacity, for the same number of drives, but achieves a high level of data integrity or fault tolerance.[3] It takes advantage of RAID Level 0 data striping methods, except that data is striped across all but one of the drives in the array. This drive is used to store parity information for maintenance of data integrity across all drives in the subsystem. The parity drive itself is divided into stripes, and each parity drive stripe is used to store parity information for the corresponding data stripes dispersed throughout the array. This method achieves high data transfer performance by reading from or writing to all of the drives in parallel or simultaneously, but retains the means to reconstruct data if a given drive fails, maintaining data integrity for the system. This concept is illustrated in Figure 5.8-7. RAID Level 3 is an excellent configuration for moving very large sequential files, such as video, in a timely manner.

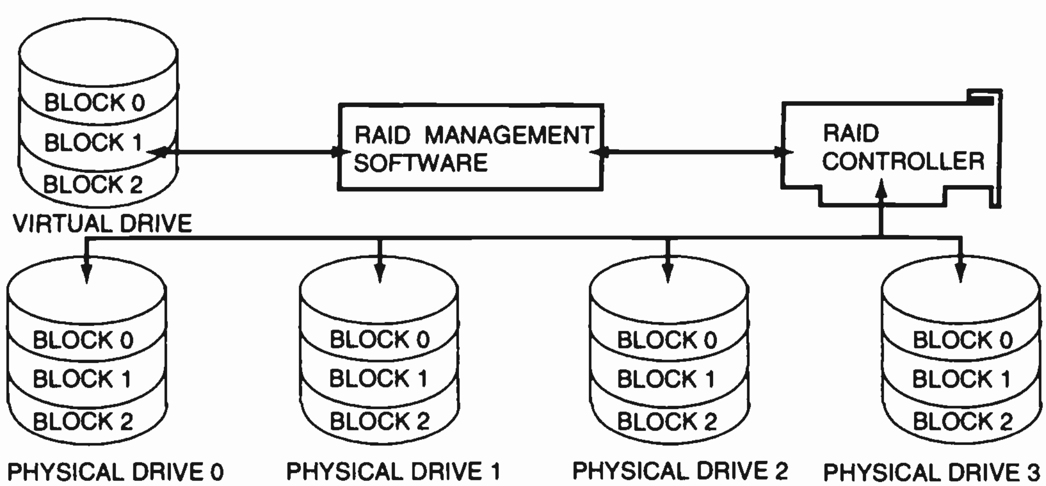


Figure 5.8-6. A RAID Level 1 subsystem.[3]

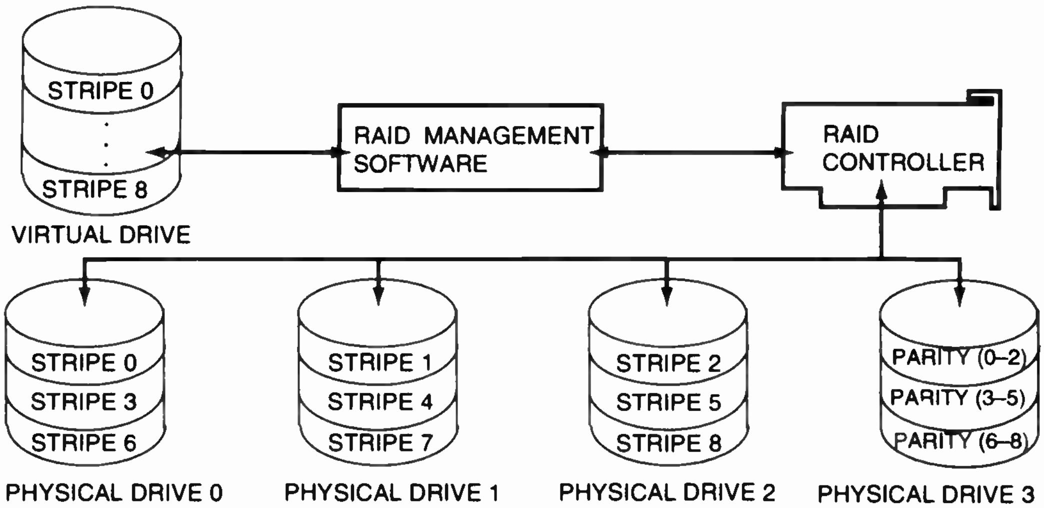


Figure 5.8-7. A RAID Level 3 configuration.[3]

The stripes of parity information stored on the dedicated drive are calculated using the *Exclusive OR* function. By using Exclusive OR with a series of data stripes in the RAID, any lost data can be recovered. Should a drive in the array fail, the missing information can be determined in a manner similar to solving for a single variable in an equation.

RAID Level 4

This level of RAID is similar in concept to RAID Level 3, but emphasizes performance for particular applications (database files versus large sequential files).[3] Another difference between the two is that RAID Level 4 has a larger stripe depth, usually of two blocks, which allows the RAID management software to operate the disks more independently than RAID Level 3 (which controls the disks in unison). This essentially replaces the high data throughput capability of RAID Level 3 with faster data access in read-intensive applications (see Figure 5.8-8).

A shortcoming of RAID level 4 is rooted in an inherent bottleneck on the parity drive. As data is written to the array, the parity encoding scheme tends to be more tedious in write activities than with other RAID topologies. This more or less relegates RAID Level 4 to read-intensive applications with little need for similar write performance. As a consequence, like Level 2, Level 4 does not see much common use in commercial applications.

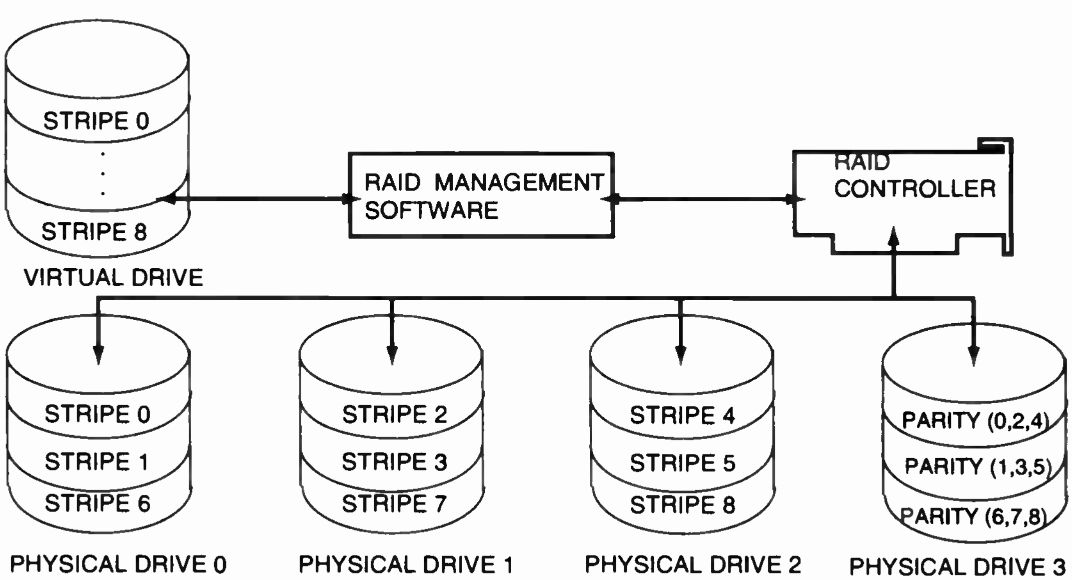


Figure 5.8-8. RAID Level 4.[3]

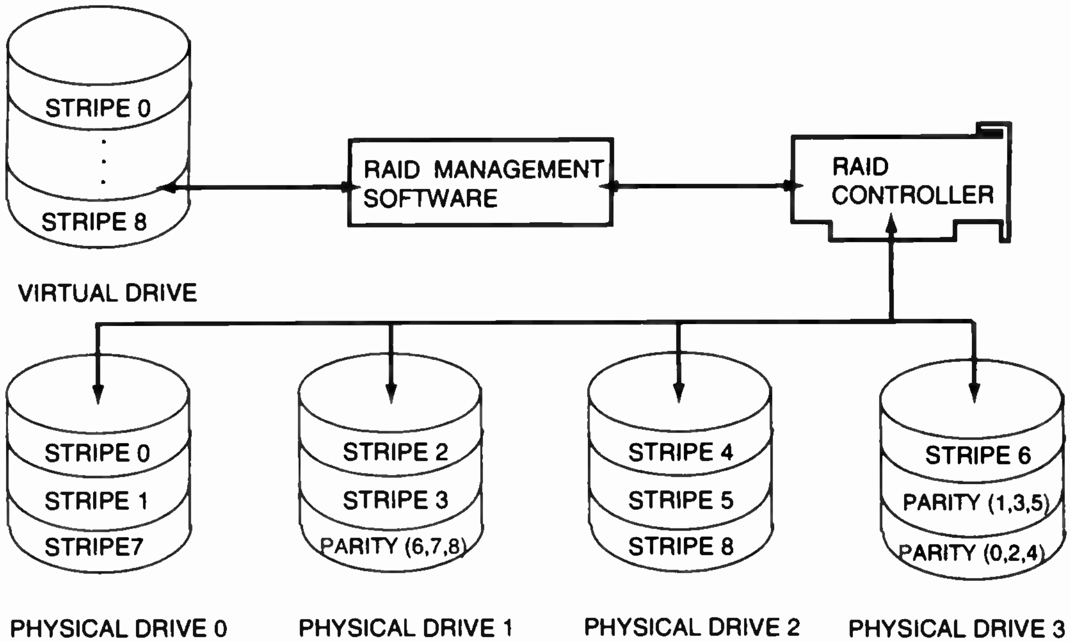


Figure 5.8-9. RAID Level 5 configuration.

RAID Level 5

Level 5 is the last of the common RAID levels in general use, and is probably the most frequently implemented.[3] RAID Level 5 minimizes the write bottlenecks of RAID Level 4 by distributing parity stripes over a series of hard drives. In so doing, it provides relief to the concentration of write activity on a single drive, which in turn enhances overall system performance (see Figure 5.8-9).

The way RAID Level 5 reduces parity write bottlenecks is relatively simple. Instead of allowing any one drive in the array to assume the risk of a bottleneck, all of the drives in the array assume write activity responsibilities. This distribution eliminates the concentration on a single drive, improving overall subsystem throughput. The RAID Level 5 parity encoding scheme is the same as Levels 3 and 4, and maintains the system's ability to recover lost data should a single drive fail. This recovery capability is possible as long as no parity stripe on an individual drive stores the information of a data stripe on the same drive. In other words, the parity information for any data stripe must always be located on a drive other than the one on which the data resides.

Other RAID Levels

Other, less common, RAID levels have been developed as custom solutions by independent vendors, including:

- RAID Level 6, which emphasizes ultra high data integrity
- RAID Level 10, which focuses on high I/O performance and very high data integrity
- RAID Level 53, which combines RAID Level 0 and 3 for uniform read and write performance.

Perhaps the greatest advantage of RAID technology is the sheer number of possible adaptations available to users and systems designers. RAID offers the ability to customize an array subsystem to the requirements of its environment and the applications demanded of it. These attributes make RAID systems an integral element in applications requiring mass storage of video signals, such as video servers. Table 5.8-4 lists the relative attributes of the various RAID implementations.

Table 5.8-4
Summary of RAID level properties.

RAID Level	Capacity	Data Availability	Data Throughput	Data Integrity
0	High	Read/Write High	High I/O Transfer Rate	
1		Read/Write High		Mirrored
2	High		High I/O Transfer Rate	ECC
3	High		High I/O Transfer Rate	Parity
4	High	Read High		Parity
5	High	Read/Write High		Parity
6		Read/Write High		Double Parity
10		Read/Write High	High I/O Transfer Rate	Mirrored
53			High I/O Transfer Rate	Parity

Digital Media Applications

Speed and storage capacity are particularly important to digital media applications (those applications that record or play audio and/or video data).[3,4] One minute of uncompressed CD-quality audio requires over 4.5 MB of storage. That is over 150 times the storage required to hold most word processing documents. The typical data rate requirement for a single channel of uncompressed digitally encoded audio is 76 KBps.

For digital video, the capacity requirements jump exponentially. A single frame of 70 mm film, if stored in uncompressed format, requires 40 MB of disk storage. Video is typically compressed to increase the amount of information that can be stored on the disk and to reduce the high data rate requirements. The data rate varies depending upon the video compression techniques used (JPEG, MPEG or other proprietary schemes) and the desired quality of the stored video image. One minute of broadcast quality video using JPEG compression requires at least 120 MB of storage. The data rate can vary from 150 KBps for MPEG compression, up to 27 MBps for full frame, uncompressed, interlaced video.

Simply put, the higher the video resolution the more space and higher data rate it requires. Where a 4 GB capacity drive might have been big enough for a mid-quality video production, higher quality video may require 9 GB of storage for a given segment.

Storing audio and video information on disk drives is a logical step on the digital media road because it provides instant access to stored data. However, not all disk storage is created equal. Optimizing disk drives for digital media requires more than just high capacity and fast data rates. Data processing applications read and write small chunks of data stored more or less randomly around the disk. Consequently, conventional disk drives have been optimized for small block, random data transfers. Yet, while contemporary data processing disk drives deliver data to traditional applications quickly, they usually have difficulties satisfying the requirements for real-time audio and video applications. Because of the nature of audio and video playback, digital media applications require that data be delivered on time, at the required rate, with no delay. Any delay in the delivery of data from the disk will cause noticeable interruption in the playback, resulting in audio breaks and/or missing video frames.

The linear nature of sound and video playback dictates that these files must be organized contiguously on the disk in much the same way as they are on tape. With tape, the media containing the information travels past the read head at the velocity required to present the sound and images at a real-time speed. In this case, the next word, musical note or video frame always resides adjacent to the current sound or picture. It is the sequential linearity of tape that ideally suits it for these purely playback applications. When the application demands manipulation or editing of the audio or video information, the utility of linear tape storage methods quickly become dubious and cumbersome.

For video servers and non-linear audio and video production systems, a storage device is required that can instantly access any video frame, sound or word and record or playback with the same-or better-measured precision as tape. This is a type of storage for which standard data processing disk drives were not designed. Figure 5.8-10 charts the speed and smoothness of a disk's throughput by graphing worst case data access over time. The performance of the tested drive is indicated by the broken line. The straight, solid line above the X axis indicates the slowest acceptable access time to deliver video data. Every time the disk drive data response time exceeds this value, the video playback will experience dropouts or lost frames, which result in visual and/or audible discontinuities in the playback.

Previously, data interruptions were solved by adding large memory buffers in the host system to hold several frames of data while the disk performed the requisite housekeeping chores. Although host memory buffers cannot be completely eliminated, their size and subsequent cost can be significantly reduced by minimizing the time the disk spends attending to certain housekeeping functions. A disk drive optimized for continuous recording and playback will maintain a virtually constant data rate to ensure that every picture and every sound is recorded and played back without any evident errors. The broken line in Figure 5.8-11 illustrates the performance of a drive optimized for audio/video applications.

Different manufacturers approach the challenges of designing drives optimized for audio/video applications in different ways. One design achieves the desired relatively flat data rate performance by changing the way disk drives work. By designing disk storage with the requirements of digital media in mind—including the development of new on-disk caching schemes optimized for the file characteristics of audio and video applications and the implementation of *intelligent housekeeping* techniques to hide necessary tasks in the background—the overall performance is improved considerably.

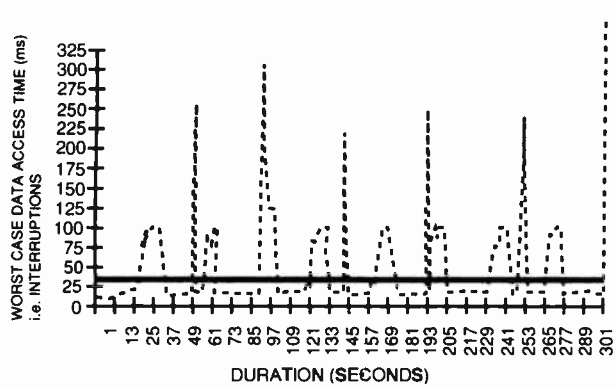


Figure 5.8-10. Disk drive data access as a function of play time for a conventional hard drive. Host demand rate = 4.0 MBps.[3]

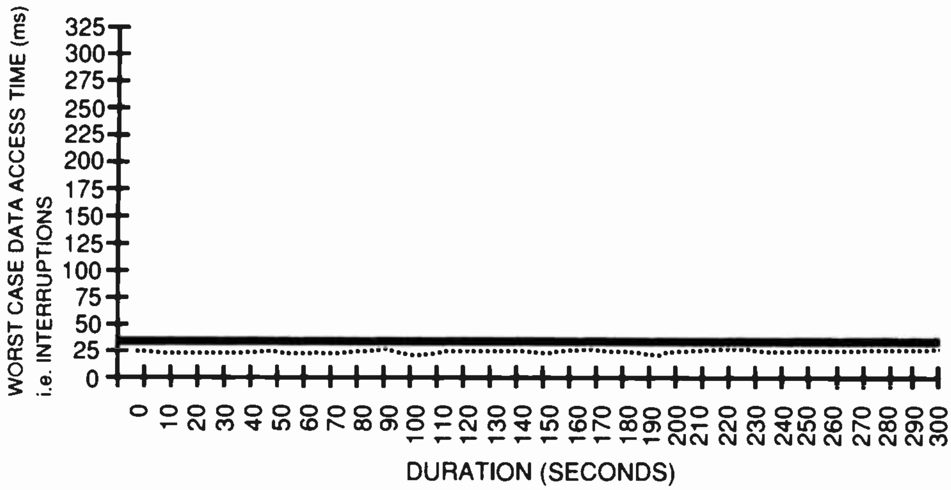


Figure 5.8-11. Disk drive data access as a function of play time for a hard drive optimized for A/V applications. Host demand rate = 4.0 MBps.[3]

Optimizing Digital Media Disk Performance

Although the error recovery methods used in conventional disk drives are adequate to keep disks operating within desired parameters, they are not optimized for speed.[3,4] To maintain the disk drive at its optimum operating range, or to correct errors, some additional issues are important. Issues that can interrupt the flow of data include the following:

- **Error recovery procedure overhead.** The error recovery procedure within most conventional disk drives is very robust and thorough. However, it is a major contributor to data flow interruptions due to soft data errors. For example, the time required for a worst case recovery from a correctable data error can often exceed 750 ms. Most digital media applications cannot tolerate a delay of this magnitude without a disruption in the audio or video playback. One solution to this situation is to implement sophisticated and exceptionally reliable means of error correction that maintains data integrity while rapidly completing housekeeping and recovery tasks. The time required to perform error recovery procedures can, in most cases, be held to 10 ms or less
- **Hardware ECC on the fly.** By adding a dedicated hardware correction engine, hardware-implemented error correcting code can be implemented *on the fly*. In other words, a soft read error can—in most cases—be corrected by a dedicated hardware correction engine. This operation is done within a few hundred microseconds, without incurring the overhead of waiting a full rotation in order to read the same data again. The more conventional approach uses a software error correcting code technique which is many times slower than the hardware ECC approach
- **Rotational retries.** If hardware ECC is not present in the drive or is unsuccessful in properly recovering the data, the next step in the error recovery process is to simply retry the operation. This approach, however, is not optimized for speed. When a disk tries to re-read data it could not retrieve on the first pass, at least one revolution is lost. The ECC on the fly techniques can usually recover soft read errors without extra disk revolutions.
- **Thermal calibration (T-Cal).** Thermal calibration is a periodic (every 10 min) housekeeping function that is necessary for disk drives incorporating have a dedicated servo system. Dedicated servo designs are commonly used by many disk drive manufacturers. The T-Cal operation calibrates the servo system to ensure that the disk heads remain precisely over the data tracks by compensating for temperature changes during normal operations. The T-Cal operation, depending upon the exact design, can require hundreds of milliseconds to complete. During all or at least part of this process, the drive will not be responsive to data requests. This is one of the major causes of dropped video frames and interrupted audio during playback of digital media. Drives optimized for audio/video applications can employ a hybrid servo system, which combines the best features of both dedicated and embedded servo system designs and eliminates the requirement for periodic T-Cal operations. Through clever design, it is possible to eliminate T-Cal operations, not simply defer them
- **Data head degaussing.** During the normal course of disk operation, the data heads accumulate a slight magnetic orientation. This residual magnetism can adversely affect operation. To prevent this condition, the drive must degauss each head. Because head degaussing involves at least one head seek plus some rotational latency, this activity can cost more than 40 ms and will interrupt the flow of data from the disk. Refined degaussing schemes minimize head seeking and rotational latency.

By combining the aforementioned techniques, it is possible to significantly improve the overall sustain-able throughput of disk drives so that they can reliably be used for demanding audio/video applications.

Thermal Calibration

It is worthwhile to explore the issue of T-Cal in greater detail. To understand the need for thermal calibration, the physics of disk drives must be examined.[5] All metals expand and contract to a certain degree under varying temperatures. The same is true with disk drive parts, including platters and the actuator arms that read the data. As a drive spins, the components heat up and expand. Although the size variance may at first seem minimal, it is potentially rather significant as tracks are spaced closer together on disk drives. Soon, the actuator arm may have moved over one or more tracks from its original course.

Initially, drives circumvented this problem by delaying read/write operations long enough to recalibrate the location of the actuator arm over the platter. This would take a few milliseconds, and would be unnoticeable in standard applications. With A/V projects, however, a pause lasting several milliseconds results in skipped frames and lost audio output.

Several approaches have been developed to solve this problem. One involves the repositioning of the heads. Normally, all heads of a disk drive would recalibrate at the same time, bringing a complete stop to the data flow. To minimize the duration of data flow interruptions, newer drives reposition one head at a time. Further, the drive can complete a thermal calibration by taking advantage of its onboard cache. A segment of data is read ahead of time and stored in the drive buffer. When that data is requested from the computer, the drive dumps it from the cache back to the system. While the buffer is being read, the actuator arms take the opportunity to recalibrate.

SCSI and Fibre Channel (FC)

Although not specifically a function of disk drive technology, new variations of SCSI and *fibre channel arbitrated-loop* (FC-AL) are important in the broad view of storage technology because of their impact on data transfer rates.[3] For this reason, they will be examined briefly here.

Since its inception in the early 1980s, SCSI has evolved to become a widely accepted and successful disk drive interface.[6] Despite its venerability, however, parallel SCSI still has its drawbacks, including cable distance limitations, a confusing mix of variants to the bus (*Fast, Fast&Wide, Differential, Ultra,* and so on) and a ceiling on the number of peripherals it can support. SCSI throughput performance has also reached a point where disk drive capacities and data rates are likely to render current SCSI variants as bottlenecks in future systems built around fast new microprocessors. These and other limitations of SCSI are creating demand for a better solution to the high-performance I/O needs of the computer systems market, FC-AL is one solution. FC-AL is a subset of the box-to-box standard created by original members of the Fibre Channel Association, Hewlett Packard, IBM and Sun Microsystems. Like its Fibre Channel superset, FC-AL is an industry standard interface endorsed by the American National Standards Institute (ANSI). Fibre Channel is usually thought of as a system-to-system or system-to-subsystem interconnection standard that uses optical cable in a point-to-point or switch configuration. This is what it was originally designed to do, in fact, HIPPI and the Internet Protocol were among the protocols defined for it. One of the goals in the development of the Fibre Channel interface was to improve or eliminate SCSI shortcomings, particularly in the areas of connectivity, performance and physical robustness.

In 1991, the Fibre Channel box-to-box interconnect standard was enhanced to include support for copper (nonoptical) media and multidrop configurations, both of which enable the low-cost connection of many devices to a host port. FC-AL made it possible for Fibre Channel to be used as a direct-disk-attachment interface (SCSI-3 has been defined as the disk interface protocol, specifically SCSI-FCP). The implications of that capability are enormous in terms of the cost savings and ease with which users will be able to migrate to standard systems with performance capabilities heretofore only found in expensive proprietary systems at the workstation or mainframe level.

The basic features of Fibre Channel are listed in Table 5.8-5. The interface features of FC-AL are given in Table 5.8-6.

FC-AL Topology

As shown in Figure 5.8-12, the FC-AL interface is a loop topology, not a bus in the conventional sense like SCSI.[3] It can have any combination of hosts and peripherals, up to a loop maximum of 126 devices. Table 5.8-5 summarizes the basic features of FC-AL.

Table 5.8-5
Fibre Channel features and capabilities.

Parameter	Range of Capabilities
Line rate	266, 531, or 1062.5 Mbps
Data transfer rate (maximum)	640–720 Mbps # 1062.5 line rate
Frame size	2112 byte payload
Protocol	SCSI, IP, ATM, SDI, HIPPI, 802.3, 802.5
Topology	Loop, switch
Data integrity	10E-12 BER
Distance	Local and campus; up to 10 km

Table 5.8-6
Basic parameters of FC-AL.[3]

Parameter	Range of Capabilities
Number of devices	126
Data rate	100 MBps (1.062 GHz using an 8B/10B code)
Cable distance	30 m between each device using copper (longer, with other cabling options)
Cable types	Backplane, twinaxial, coaxial, optical
Fault tolerance	Dual porting, hot plugging

Using a connector based on the 80-pin parallel SCSI *single connector attachment* (SCA), Fibre Channel disk drives attach directly to a backplane. This not only eliminates cable congestion, it makes hot drive insertions practical and simplifies mechanical designs. Fewer cables and components translate to lower-cost systems and higher reliability.

Figure 5.8-13 shows the use of a backplane with dual-ported drives and dual loops. This would be typical of a disk array subsystem with hot-plug capability. The drives are attached directly to the backplane for both signal and power; there are no other connections. No jumpers, terminators or switches need be set on the drive. An FC-AL drive can be simply plugged into the backplane, and the Fibre Channel controller determines the drive address from either the relative position on the backplane or the drive's unique IEEE Fibre Channel address (all Fibre Channel interface devices have a worldwide unique address). The backplane and drives can be located in the same cabinet as the rest of a computer system, or connected as the separate units shown in Figure 5.8-13. Each of the loops in this system would be capable of transferring data at 100 MBps. Note that the two systems are on different loops for redundancy. Both computers could just as easily be on both loops for full fault tolerance, or share a single loop for the sake of economy.

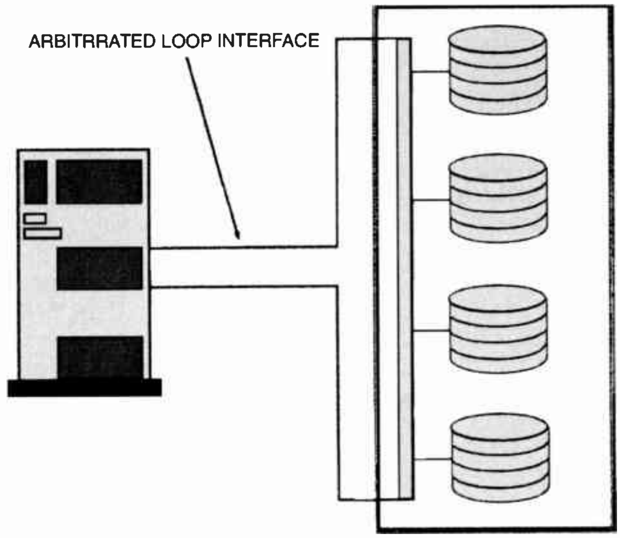


Figure 5.8-12. Basic topology of FC-AL.[3]

Connectivity Considerations

Applications such as video and image processing have pushed the demand for huge increases in disk capacity per system.[3] In some cases, the capacity requirements for these types of applications are such that it is difficult to configure a sufficient number of SCSI

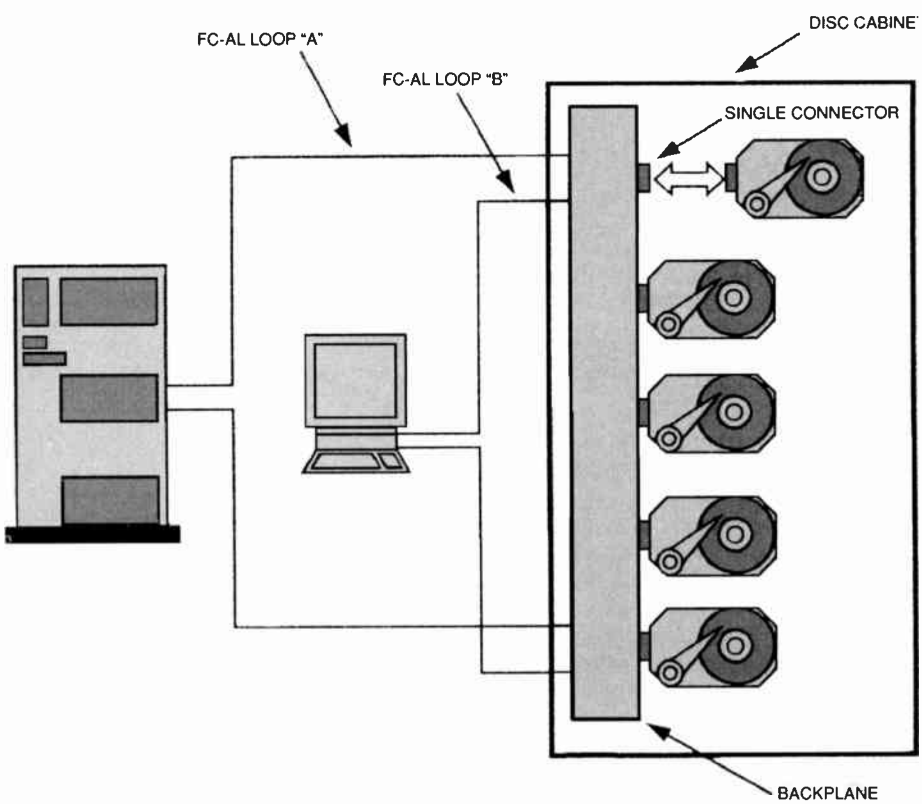


Figure 5.8-13. An array of disk drives in a dual loop FC-AL system.[3]

buses on a system so that enough drive addresses are available to attach the necessary disk storage. Moreover, simply increasing the addressability of SCSI—making it possible to have more than 15 devices per wide SCSI bus, for example—would not be a solution, because more bus bandwidth is needed to support the additional drives. Besides, protocol overhead is already excessive. FC-AL can address up to 126 devices, but practical usage is another matter. A loop can practically support about 60 drives in a UNIX (with 8 KB I/Os) environment, as an example. With a 9 GB FC-AL interface disk drive, a loop of 50 drives would make more than 450 GB available on a single FC-AL host adapter. This makes it possible for any workstation or system with a Fibre Channel port to become an incredibly large file server. When the cost of a single host adapter is compared to the many channels and controllers that a mainframe would employ to attach the equivalent capacity, the economics of Fibre Channel attached-disk storage become obvious.

Bandwidth

A Fibre Channel loop, as stated previously, supports data rates up to 100 MBps in single-port applications and up to 200 MBps in dual-port configurations.[3] Applications such as digital video data storage and retrieval, computer modeling and image processing are growing in popularity and demanding ever increasing improvements to disk data transfer performance. Moreover, file servers are increasingly looked upon as replacements for mainframe computers. In order to fulfill that promise, they will need to deliver higher transaction rates to provide a mainframe level of service.

Magnetic disk drive areal densities are known to be increasing at about 60% per year in production products. Because the number of bits per inch—one of the two components of disk areal density—is increasing at about 30% per year, the data rate performance increases proportionately. In addition, drive spindle speeds (or spin rates) have more than doubled over the last five years, and any spin rate improvements will also directly contribute to improved data rate performance.

Remote Online Storage

Because the FC-AL interface is part of, and fully compatible with the Fibre Channel Standard, optical cabling can be used in any portion of a subsystem (with the exception that optical signals, of course, cannot be used on a backplane).[3] This makes it possible to have a disk subsystem a significant distance from the computer system to which it is attached. For example, using single-mode fibre optics, online disk storage could be as far as 10 km away from its host. As illustrated in Figure 5.8-14, the computer system on the left could have disks within the system connected via a Fibre Channel loop. That loop can be extended by using an electronic-to-optic signal adapter and lengths of fibre optic cable. On the same loop running the internal disks, remote disks would appear to the system to be directly attached exactly like the local disks, even though they might be five miles away. This can be an advantage in many ways, including the shadowing the local disks in the event a disaster destroys the on-site data banks. That capability, in particular, offers an attractive means for having a remote and secure online copy of critical data that could be used to continue operations should anything happen to the facility housing the primary computer system.

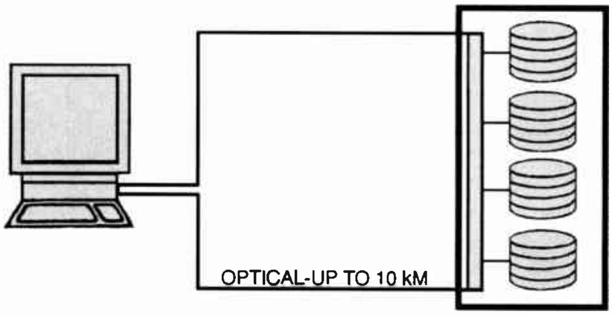


Figure 5.8-14. The use of optical interconnection for long-distance FC-AL service.[3]

Array Implementations

Array controllers have traditionally been designed with multiple SCSI interfaces for drive attachment.[3] This enables the controller to supply data and I/O rates equal to several times those achievable from a single interface. This is sometimes referred to as an *orthogonal array*, because the disks comprising the arrays elements, are across, or orthogonal to, the SCSI controllers. The decision to design a specific number of drive interfaces into a given controller forces on the customer the parity amortization, granularity and controller cost associated with that decision. This limits the customer's choices for configuring the optimal combination of economy—that is, maximizing the *granularity*, or number of data drives per parity drive, the total capacity per array and overall performance. In the example shown in Figure 5.8-15, the array has six rows, amortizes the parity data over five data drives, and often requires adding six drives each time any increase in capacity is required. With Fibre Channel, it is possible to configure an array along a single interface instead of across many (see Figure 5.8-16). Because the drives constituting the array unit are organized along an interface, it is sometimes described as a *longitudinal array*. While a single loop makes a perfectly legitimate array configuration, the diagram shows an optional second loop and controller for fault tolerance.

Some of the more important benefits from using FC as the basis of a longitudinal array include:

- A system designer has the flexibility to decide the economy of the array. It can utilize 5, 8, 18 or 24 data drives per parity drive, with no additional controller cost
- The granularity is on a drive-by-drive basis—the system designer is not required to add a whole row

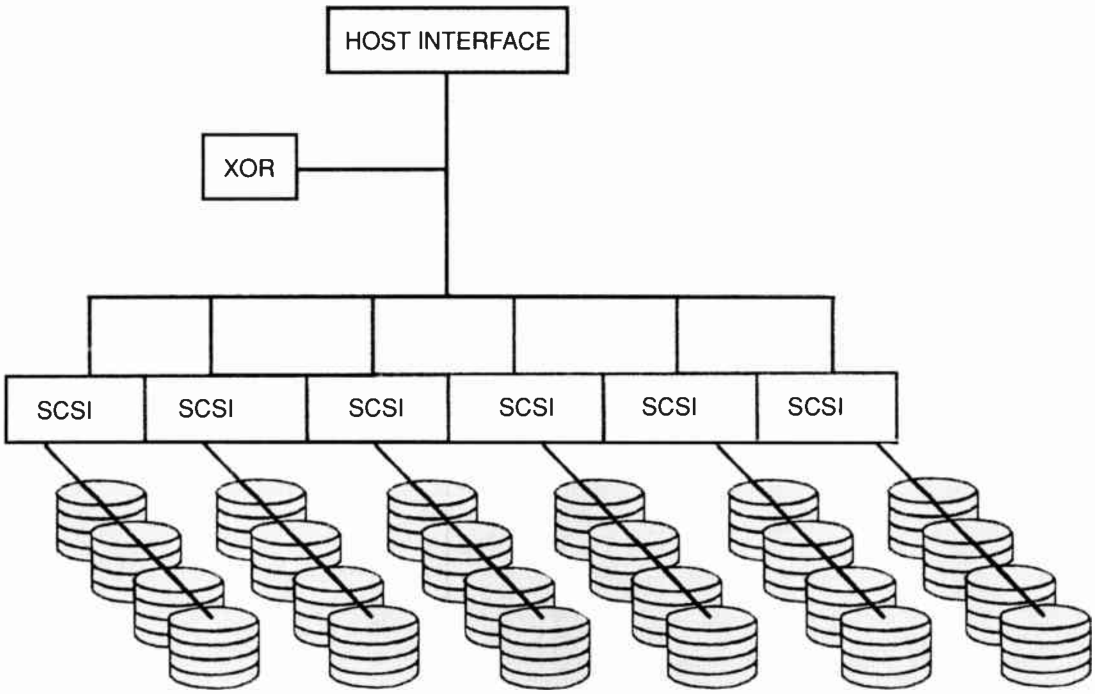


Figure 5.8-15. A conventional SCSI orthogonal array.[3]

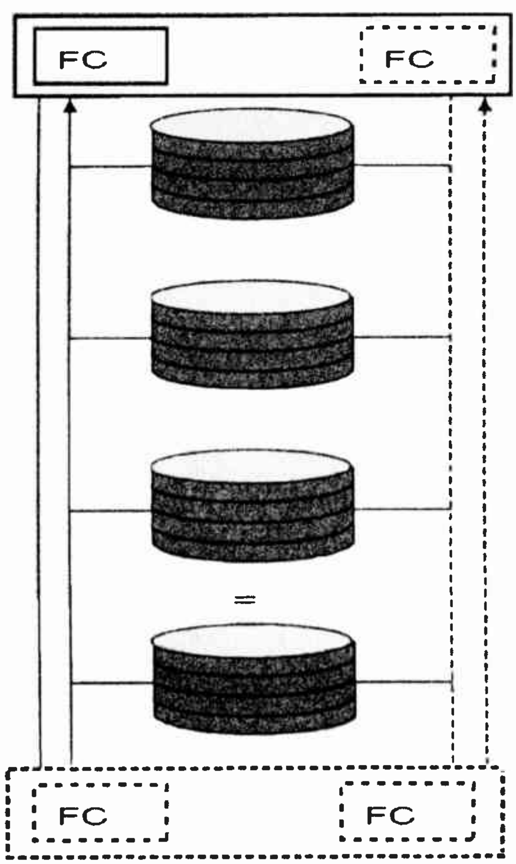


Figure 5.8-16. A Fibre Channel longitudinal array.[3]

of drives as on a traditional array. The capacity of a subsystem can be increased on a drive-by-drive increment basis

- The cost of the array over conventional array implementations is smaller. Because the FC-AL controller has been significantly simplified, it is less expensive than a multiple drive interface version
- A designer can add controllers and spread drives across them as more performance is needed (but only add drives if extra capacity is desired). The designer can scale performance and cost per megabyte to best suit a particular application.

Networking Considerations

Because Fibre Channel is a generic, standard interface, it has multiple uses and supports many different protocols, such as SCSI, IPI-3, Internet Protocol and asynchronous transfer mode (ATM).[3] All of these can run on the same FC facility, which means that an investment in a FC adapter for a disk array may also bring high performance networking and vice versa.

While relatively new to the scene, *network-attached storage* offers intriguing possibilities and Fibre Channel makes it architecturally practical. Figure 5.8-17 shows an FC network, including a file server with an arbitrated loop of drives. Because the disk interface is the same as the network interface, there is no reason the disks could not be put directly on the network, as with the two drives depicted in the upper left section of the loop. The systems could be communicating with each other using IP protocol and with the disks using some variation of SCSI over the same facility. All the

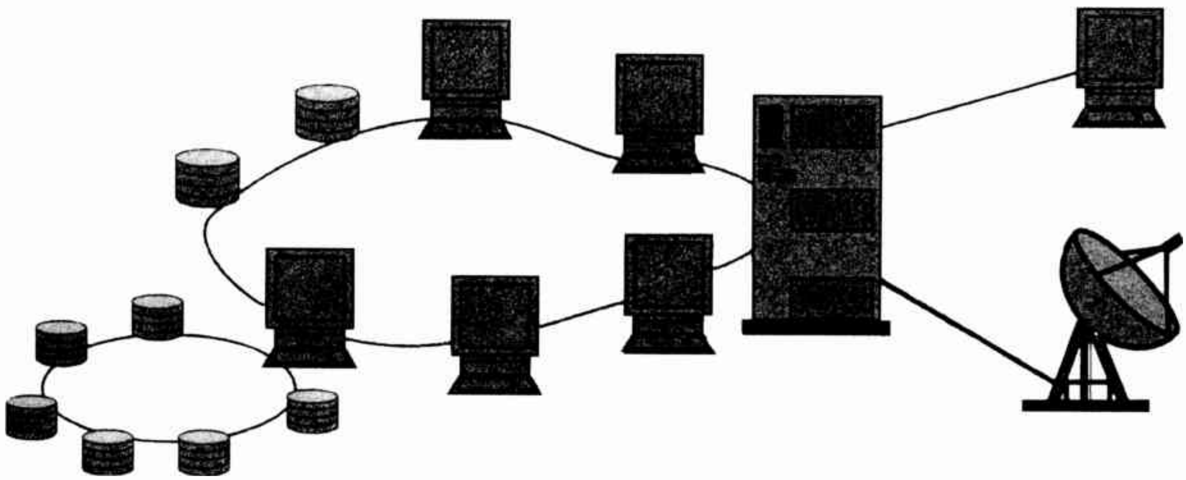


Figure 5.8-17. The concept of network-attached-storage.[3]

systems in the network could access that data at Fibre Channel speeds. This might be especially attractive in a video server application where the data could be sent directly from the drive to a cable head end or satellite dish at speeds much higher than could be achieved if routed through a file server.

Server-Based Video Editing

Video editing is mostly a cutting and pasting process, and as such, it lends itself to synergistic applications involving the video server.[6] The digital recorder enabled users to re-record the same video several times with minimal generation loss. Digital recording also offered the ability to read-before-write. With tape, this required judicious care, because the underlying video track was erased by the subsequent one. Disk based systems, either optical or magnetic hard disk drive designs, also offer this capability, but allow nondestructive read-out.

Server based editing resembles digital disk based editing, except that files from one user can be instantly available to another. Facilities need only to endure the time penalty of transfer and digitizing (as necessary) the input source material once. Thereafter, all potential users can access the material simultaneously. For example, assume an important piece for the five o'clock news is being produced in edit room "A." Via the server, the producer in edit room "B" can start putting together the same story for the six o'clock news, accessing the same digitized elements.

Not every facility puts out back-to-back newscasts. Many that do, reuse stories with minimal updating. There are, however, a number of facilities that not only produce multiple newscasts, but provide separate news programming to cable channels or that sell news to other stations. If Internet broadcasting increases in importance, it may begin to consume editing resources as well. With this many hands fighting for a field tape, conflict is inevitable. A unique economic advantage of the server based facility is that a multitude of users can cherry pick off the main storage system, without disrupting the work flow of other operation.

CONCLUSION

The video industry has become accustomed to an ever increasing number of tape formats, each targeting a particular market segment. Disk storage of video has been an effective refuge from such format wars. This is not to say that video disk stores do not vary greatly—they do. The only standardization is their interface: SCSI, SCSI-1, SCSI Wide, Fiber Channel, IDE and so on. Because video disk interchange is not a major requirement today, there is no specific need to define the internal format. With no standards of measurement except the ultimate performance of the system to fall back upon, it is important to carefully examine the system architecture to be certain that it addresses the requirements for performance and reliability for a given application. Clearly, all video servers are not created equal.

The utilization of server systems opens considerable opportunities for video production. The ability to operate multiple channels, provide multiple, simultaneous access, reduce the space required per task and better track program materials is helping to define the video production center of the future.

REFERENCES

1. Robin, Michael, and Michel Poulin, "Multimedia and Television," *Digital Television Fundamentals*, McGraw-Hill, New York, pp. 455–488, 1997.
2. McConathy, Charles F., "A Digital Video Disk Array Primer," *SMPTE Journal*, SMPTE, New York, N.Y., pp. 220–223, April 1998.

3. Whitaker, Jerry C., "Data Storage Systems," *The Electronics Handbook,* Jerry C. Whitaker (ed.), CRC Press, Boca Raton, FL, pp. 1445–1459, 1996.
4. Portions of this section based on background information provided by Micropolis, Chatsworth, CA, 1996.
5. Tyson, H., "Barracuda and Elite: Disk Drive Storage for Professional Audio/Video," *Seagate Technology Paper #SV-25,* Seagate, Scotts Valley, CA, 1995.
6. Lehtinen, Rick, "Editing Systems," *Broadcast Engineering,* Intertec Publishing, Overland Park. KS, pp. 26–36, May 1996.

BIBLIOGRAPHY

Anderson, D. "Fibre Channel-Arbitrated Loop: The Preferred Path to Higher I/O Performance, Flexibility in Design," *Seagate Technology Paper #MN-24,* Seagate, Scotts Valley, CA, 1995.

Goldberg, Thomas, "New Storage Technology," *Proceedings of the Advanced Television Summit,* Intertec Publishing, Overland Park, KS, 1996.

Heyn, T., "The RAID Advantage," *Seagate Technology Paper,* Seagate, Scotts Valley, CA, 1995.

Plank, Bob, "Video Disk and Server Operation," *International Broadcast Engineer,* September 1995.

5.9 VIDEO SPECIAL EFFECTS SYSTEMS

JEFF MAZUR, ABC INC.
HOLLYWOOD, CA

INTRODUCTION

Special visual effects is one of the fastest growing fields in entertainment production. The proliferation of music videos, commercials and even video games has created an insatiable need for newer and better effects. Once used only for truly special effects, or impossible shots (*Bewitched* style magic or sci-fi environments such as outer space) they are now routinely used to make subtle changes to brand names, lighting and to create realistic scenes such as the drifting feather in the movie *Forest Gump*. Although more prevalent in large budget motion pictures, most of these effects are now being performed digitally in a high resolution video environment. Therefore, the techniques described here apply equally well to standard and high definition television as well as to digital film effects equipment.

There are many techniques that go into the making of special effects, from mechanical (wires) and chemical (pyrotechnics) to motion control, camera tracking and virtual sets. Some of these effects are created in *real time*, during original camera shooting. Other visual effects such as 2D and 3D computer generated images (CGI) and digital effects are added later during post-production. This chapter focuses primarily on the use of digital video effects (DVE) devices and compositing systems.

History

The history of digital video effects dates back to the early days of television when the only effect that could be performed was superimposition (mixing two video feeds) similar to making a double exposure on film. Normally used to add graphics or credit rolls to a live video feed, the video mixer was simply dissolved halfway between two cameras. Although the effect was crude—and lowered the brightness level from each camera—it was possible to create composite images of ghosts, giants, etc. For this to work, one subject had to be shot against a black background to avoid spoiling the other image.

Later, video switchers developed the ability to actively switch between two cameras thus creating a wipe. With the advent of the videotape recorder (VTR) and careful planning, it became possible to wipe between a live camera and tape playback, allowing one actor to play against herself, much like the traditional optical effect used in film.

Analog Effects

A *keyer* enables the switching between two video sources. The switching signal is derived from a video source rather than a fixed pattern generator. An internal key typically uses the luminance level of the video to create the switch; this is practical for superimposing black-and-white graphics or text.

For effects work, the next advance took place with the introduction of the *chroma keyer*. A chroma keyer generates a key signal based upon the color of objects in the scene. Thus an actor posed in front of a strongly lit blue or green screen can be isolated from that picture and placed over, or into, a completely different scene from another camera.

For more than 20 years, this was about the extent of the effects that could be performed on video signals. In the early 1970's however, a revolution was taking place in the field of electronics. In the broadcast arena, it began with the digital time base corrector (TBC). By storing a small portion of the video signal in random access memory (RAM), a digital TBC could eliminate jitter from the raw output of a VTR. As memory capacities grew, it was soon feasible to hold an entire frame of video in RAM, thus creating the *frame store*. At this point, many inventive engineers realized the potential this had for creating digital video effects.

DIGITAL VIDEO EFFECTS SYSTEMS

The techniques of sampling, digitizing, RAM storage and conversion from digital to analog video were honed on the TBC and frame store. With an entire frame of memory, video could be stored into RAM at one rate and read out at another. This is the basic operation of the frame store—allowing asynchronous video signals to be combined without the need for genlock. In this scenario, the digitized video samples from the incoming signal are stored more or less sequentially into the RAM buffer. They are also read out sequentially although the offset between the read and write addresses will change depending upon the difference in frequency and phase of the incoming video versus the station reference.

However, it should be obvious that there is no restriction on how the video samples can be retrieved

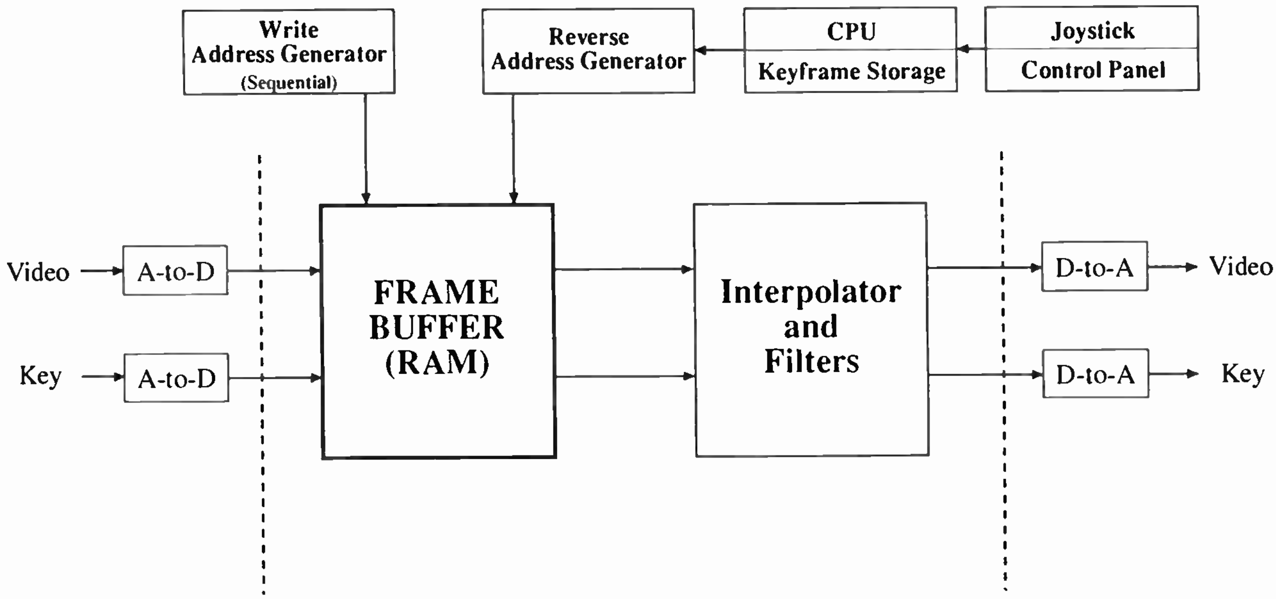


Figure 5.9-1. Block diagram of a typical standalone DVE device. Most effects are created by the Reverse Address Generator (RAG), under control of the CPU.

from memory. For example, start with the assumption that each digital sample represents one pixel of video.[1] If the read addressing is then altered so that the video pixels are read out from RAM in reverse order, there would then be an upside-down and mirror image picture coming out of the framestore (a digital video effect).

When trying to define the various effects capabilities of a DVE, it is helpful to think in terms of source and destination spaces. Imagine the original video signal which is fed into the DVE as represented by a flat piece of paper; that would represent the two dimensional source space. Since the output is also a video signal that might be displayed on a monitor, it too can be represented by the two dimensional face of the monitor screen; this is the destination space. The relationship between these two spaces can be altered in various ways within the DVE including separation into an imaginary third dimension going into the screen. These operations are usually performed using a 3-axis joystick.

Figure 5.9-1 shows the basic block diagram of a typical DVE. First, all analog video and/or key signals are digitized by an A/D converter. The video samples are written directly into the frame buffer. The Reverse Address Generator (RAG) then determines how these video samples will be read out. The RAG is controlled by a CPU which calculates the sequences necessary to give the desired effect. The output from memory is passed through an interpolator and one or more filters to create the final digital output frame. If necessary, the output is converted back to analog. This process continues for each frame of video passing through the DVE; thus the CPU must be fast enough to calculate the RAG sequences in real time.

Translation

The simplest effect that can be performed involves *translation*, or movement in the horizontal and/or vertical direction. It easy to see that these effects can be accomplished by altering the sequence in which the stored pixels are read out. If the sequence itself is then manipulated in real time, it's possible to have the DVE smoothly move the picture left or right, up or down. When trying to read out past the edges of the source space, the DVE will usually substitute black or a fixed background color.

Interpolation

The previous examples did not alter the original video samples. They were simply read out in a different order. Almost all effects however will require some manipulation of the stored samples to create the pixels in the output image. This is the job of an *interpolator*. For each output pixel, the interpolator examines the surrounding pixels stored in memory. Then using a precisely weighted average, it calculates (or interpolates) what the final output pixel should be.

Because the NTSC video signal is an interlaced format, finding an adjacent pixel is somewhat ambiguous, at least in the vertical direction. For example, consider what would be called the pixel directly above. If the pixel from the previous line (within the same field) is used, then this ignores half of the vertical resolution contained in the other field. Looking further back into memory to the previous field, gives the true

[1] In reality, the luma and chroma data from each pixel are stored and processed separately through different sections of the DVE.

pixel which is on the line directly above. This restores the full vertical resolution but introduces another potential problem—motion artifacts. If the pixel in question represents a part of the scene that is in motion, then looking back at the previous field represents what the pixel looked like 1/60 of a second ago.

In a nutshell, if the video stream is processed as fields, then half of the vertical resolution is given up but fluid motion is retained; with frame processing, full resolution is retained at the expense of possible jittery motion or *judder*. For this reason, the frame mode is usually used only for still, high resolution images such as computer generated graphics. For most other video sources which contain any amount of movement, the field mode will usually produce better results. If the motion is great enough, it will mask much of the loss of resolution.

Many interpolators also have an *adaptive* mode which can select the optimum processing for each pixel of the video image. By comparing every pixel with its counterpart in the previous frame, the interpolator can determine whether there is motion occurring in that particular spot on the screen. If motion is detected, then that pixel is processed using field interpolation. If not in motion, then frame interpolation is used for that pixel. Although the motion detector is sometimes fooled, this mode usually produces excellent results.

Rotation and Size Manipulation

Although slightly more complicated, rotation of an image is accomplished in a way similar to translation. The read addresses are calculated using standard polar to Cartesian coordinate transforms. The center, or axis, about which an image is rotated can also be easily altered.

Size manipulation is likewise performed by mathematical operations. Shrinking a picture involves tossing away information (a one-quarter size image can be created by eliminating every other pixel and line). But to blow an image up greater than full size requires more information than is available in the original video signal. Thus the interpolator is relied upon to create the missing pixels. At best, the image will appear slightly fuzzy and at large magnifications the loss of resolution may not be acceptable.

Digital Optical Effects

If the computer generating the read addressing is fast enough, then more complicated functions can be performed. Generally referred to as *warps*, these effects look like page turns, rolls, ripples, or wrapping the video onto a solid object such as a sphere.

It is also quite an easy trick to manipulate the pixel size to give a mosaic look, add multiple images on the screen, or add video effects such as a drop shadow or glow around an image. To further enhance the illusion that a video image has been moved through three dimensional space, a DVE can add perspective size and defocus to the video or even bounce an imaginary light source off of the source video plane.

Latest Computer Effects

Going even further, computerized alteration of the video can produce morphs (changing one image into another), distort (painting the video on a sheet of rubber and then deforming it) and painterly effects (changing the video to look like an oil painting, for example). Because of the mathematical processing necessary to produce these effects, they currently (1998) cannot be created in real time. Thus they probably will not be found on a dedicated DVE device such as those used for live broadcasts.

For effects work done in post-production, non-real time compositing workstations with built-in DVE functions are used to create such effects. Although processing times vary, a typical workstation can render frames somewhere between several frames per second to several seconds per frame.

Recursive Effects

Other interesting effects can be generated by feeding all or part of the processed video back through the DVE. Much like pointing a camera at its video monitor to produce video feedback, the DVE can recycle its output video to provide recursive functions. The most popular uses for this mode are to create decays or trails from moving parts of the image.

Key Channel

Although the primary function of a DVE is to manipulate the video coming into it, most DVEs also keep track of the original source rectangle or an external input key through a separate *key channel*. This channel undergoes all of the transformations along with the video so that it can be used to key the output from the DVE onto another video signal. The key channel can also be used to create a drop shadow of the processed video onto the background.

Motion Keyframing

Sometimes a DVE is used to simply perform a single, stationary effect onto an incoming video stream. The real power of the DVE however comes with its ability to alter its effect over time. A simple effect for example, might be to move a shrunken source image from offscreen left, through the target space and then completely offscreen to the right. It would also be necessary to specify the length of time that this move should take place.

To set up such an effect, *keyframes* are created where absolute parameters are to be entered. In this case, it would be necessary to specify the picture size parameter and then create a starting and ending keyframe. These keyframes would then be set to the beginning and ending x-coordinates respectively. The DVE will then calculate the x-coordinate for all of the other frames of the effect—this is called *in-betweening*. Most DVEs have a timeline display that shows the location of each keyframe within the total length of the complete effect.

Types of Motion

Adding the fourth dimension of time to the DVE also allows manipulation within this new dimension.

(a) (b) (c) (d)

Figure 5.9-2. Various motion paths that can be applied to an effect. In all of these figures, the same eight frame video clip is moved from left to right across the screen (an integrate function has been applied to show all eight frames in the same figure). In a), a *linear* motion gives rise to equidistant spacing and thus constant velocity; in b), the effect is given an s-linear motion which starts off slower, accelerates through the move and then decelerates at the end; by varying TCB values, many other motion paths are possible such as c), constant acceleration or d), deceleration in the middle of the effect.

The simplest type of motion between keyframes is *linear motion*, where each frame moves the same fraction of the difference between the keyframes. Thus the video moves with constant velocity, in a rather mechanical fashion. Another type of motion is *S-linear*, which adds a small amount of acceleration at the start of the move and deceleration at the end. This usually gives a more natural feel to the motion. Other variations such as smooth or curve motions also try to mimic more closely the movement of actual objects in nature. Very specific motion paths, such as that of a bouncing ball, may require many keyframes or finer control over the motion parameters.

One scheme for setting precise control over keyframe motion is the use of *motion vectors*. Three parameters—tension, continuity and bias—determine how the video moves into and out of each keyframe. Figure 5.9-2 shows the effects of various motion types on a simple video clip which is moved across the screen.

For the ultimate in motion keyframing, some systems present a graph for each parameter that changes. The curve representing how the parameter changes from one keyframe to the next can then be manipulated to give the exact motion desired.

COMPOSITING

While some visual effects can be added during production (and indeed, *must* be for a live broadcast) most effects work today is done during post-production. Here, at a much more leisurely pace, the full force of the post-production tools can be focused on perfecting each effects shot. Many devices and techniques can be combined to achieve the desired look. In most cases, the final effect will come together in a compositing suite where all of the various elements are combined. A typical effects shot might consist of 5 to 10 layers: an original background plate with one or more keyed in elements, plus shadow passes, reflections, highlights or glints and possibly a CGI or hand animated element.

In its simplest form, *compositing* involves the layering, or keying, of one video image over another. This multiple image is then recorded off to tape or disk where it can be recalled later as the background image for the next layer. Thus an entire image or video clip is built up one layer at a time. In a linear compositing suite, clips would be laid off to a component digital VTR or digital disk recorder (DDR).

Today however, falling prices and increased capacities have made dedicated compositing workstations much more popular. A complete edit suite can now be built into a single box with non-linear editing, keying and digital video effects all being controlled by a single operator. The increased efficiency and ease of operation now make them the logical choice for most effects compositing.

With early compositing suites, changing something on the first or second *bed* in a 30 layer composite was very time consuming. It could take almost as long to rebuild the composite just to make a small change to one of the first layers. It was like peeling off the layers of an onion to make a change inside and then carefully reassembling the pieces back together. Modern compositing workstations now allow the operator to work with many layers concurrently, making changes to any layer at will. Other features, such as the ability to marry a matte channel to a video clip, can drastically improve the efficiency of a compositing system.

When compositing, there are basically two ways in which video sources can be combined: *mixing* or *keying*. Mixing can be performed in either an additive or non-additive fashion. Keying—also referred to as matting—can be performed using one of the existing sources as a key signal (internal key) or using a third, separate source for the key signal or matte.

Additive versus Non-Additive Mixing

When mixing two video signals, the usual operation is to add a portion of each signal such that the total percentages add up to 100. This is referred to as an *additive* mix. This duplicates the function of the dissolve lever on a switcher. Inside the switcher or compositing computer, the mix is performed by mathematically adding the two signals in this fashion:

$$V_o = M * V_1 + (100 - M) * V_2$$

where V_o is the video output signal from the mixer, V_1 and V_2 are the video sources and M is the mixing percentage.

For certain operations however, it may be desirable to combine the signals in such a way that the output video represents the level from the video source with the higher luminance. This is called a *non-additive* mix and can be used, for example, to combine mattes from multiple objects to make a single combination matte. Another common use is to add one or more garbage mattes to eliminate areas where the camera shoots past the area of interest. Mathematically, the non-additive mix can be expressed as:

$$V_o = V_1 \quad \text{where } V_1 \geq V_2$$
$$V_o = V_2 \quad \text{where } V_1 < V_2$$

Figure 5.9-3 shows the difference between an additive and non-additive mix using two *hicon* (high contrast, 0-100 IRE) matte signals.

Keying

As previously mentioned, keying involves the combination of two video sources by selectively switching between them. A separate, monochrome key signal is used to determine when to switch. A simple keyer switches between the two sources based on the level of the key signal in relation to key level and/or clip controls. The key signal can be derived from either the overall brightness level (luminance key) or from the color hue information (chroma key), or possibly a combination of both.

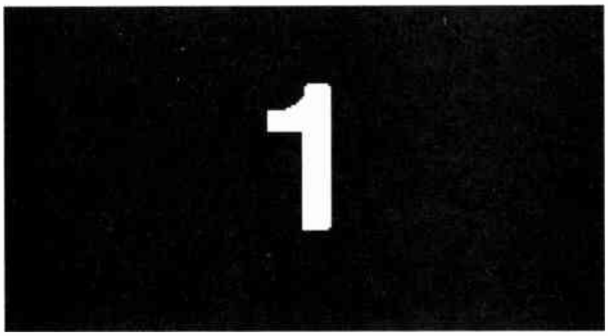
The output from a keyer can be expressed mathematically by the formula:

$$V_o = V_f \quad \text{for } K \geq K_{th}$$
$$V_o = V_b \quad \text{for } K < K_{th}$$

where V_f and V_b are the foreground and background video signals respectively, K is the key signal level and K_{th} is the key threshold level set by the operator.

The previous statement describes the operation of a hard clip keyer which switches instantaneously between video sources. Such a switch often creates undesirable artifacts which are visible in the output video. To reduce this problem, most keyers can set two thresholds over which a linear dissolve between signals is produced. This is often referred to as *softening* the key. However, the most precise control can be achieved by using a linear key.

The linear keyer, with its full range of transparency has allowed much more natural and pleasing composit-



(a)

(b)

(c)

(d)

Figure 5.9-3. Additive vs. non-additive mixing. The two hicons shown in a) and b) are combined in c), a 50% additive mix and d), a 100% non-additive mix.

ing of images. With a linear keyer, the key signal is used to effectively dissolve between the two video sources. Where the key signal is zero, or black, the foreground image will be totally transparent and thus does not show through over the background. Where the key is 100 units, or white, the foreground image becomes completely opaque and thus covers up the background. In between, as the key signal takes on other values, the foreground becomes as opaque as the absolute value of the key signal and the background shows through accordingly. This is expressed mathematically as:

$$V_o = K' * V_f + (1 - K') * V_b$$

where the key signal K is normalized to vary between 0 and 1 ($K' = K/100$).

One of the great advantages of a linear keyer is that it can maintain the proper levels of anti-aliased images (especially graphics) when creating the composite.

Chroma Key

In a chroma keyer, the key signal is extracted from the foreground video source. Thus the quality of the key will depend upon the video source as well as the chroma keyer's ability to discern colors. For a live production, the RGB outputs from a camera would be used; for later use in post-production, the blue (or green) screen shoot must be recorded. Although it is possible to pull a key from composite video, this should only be done as a last resort. All blue screen source work should be shot on 35 mm film and/or recorded directly onto a component VTR such as D1, D5, DCT, or DigiBeta. This will preserve the greatest bandwidth of the color signals and eliminate artifacts that cannot be removed from the NTSC composite signal.

Most chroma keyers work in the RGB domain. Thus, the foreground video signal must first be transformed into an RGB signal. The keyer will then analyze the individual red, green and blue levels and isolate a small range of color values to represent the key signal. It can take a lot of patience to set up the keyer properly, especially if there are other production or lighting problems. A bad key is quite noticeable; there may be parts of the foreground subject that start to disappear, or often there will be a blue halo around the keyed subject.

Edge effects around the key are the number one giveaway that a chroma key has been used. Unfortunately, even if everything is set up correctly, there may still be a colored halo around the foreground subject. Most often, this is caused by blue light from the backing material reflecting back onto the subject. It can also be caused by reflections or flares within the camera or camera lens itself. To reduce these effects, many keyers have the ability to *hue suppress* the foreground image before keying it over the background. Hue suppression selectively eliminates the backing color from the foreground image. At worst, any remaining halo would be monochrome and would blend into the background more naturally.

For the ultimate in chroma keying, an additive mixing process can be used. First the foreground image (against the colored screen) is processed to produce both a matte signal and a hue suppressed foreground image (against black). Then the two signals are combined in a fully additive mix according to the equation:

$$V_o = V_f' + (1 - K) * V_b$$

where V_f' is the preprocessed (hue suppressed) foreground signal.

Note that there is no attenuation of the foreground video, other than removal of the backing color. With this technique, very high quality keys can be achieved despite the usual problem areas such as fine details (wisps of hair) and transparent objects such as glass. Even reflections from shiny surfaces and shadows cast onto the backing wall can be recreated in the final composite.

Effective Chroma Keying

Since many effects rely upon chroma keying, it is worth discussing the various parameters that go into making a believable composite. Blue (or green) screen effects can basically be divided into two categories, depending upon whether the foreground and background cameras move or not. Early chroma keys were mostly performed with locked down camera shots. Since the purpose of the chroma key is to place a new object into an existing background scene, there could be no relative motion between the two without destroying the illusion. Later advancements with motion control rigs, motion tracking and virtual set technologies have allowed moving chroma keys to be created. Since most people are used to seeing such effects shots as stationary, these new effects seem even more believable.

Whether stationary or moving, there are many tricks that can be used to make a successful chroma key. First, there are some decisions which need to be made during pre-production. This would include the choice of backing color and the type of keying required. Knowledge of the keyer to be used for compositing and whether the original shoot will be captured on film or video will aid in the effect's planning.

Backing Color

Although technically a chroma key can be pulled from any solid background color, blue and green screens are used almost exclusively. In many cases, the choice between the two is very subtle and may be dictated solely on the availability of a suitable stage. When given the option however, there are several factors to consider. The most common concern is with colors that are to be used in the foreground scene. Conventional wisdom implies that there shouldn't be any strong blues on a subject that is to be shot against a blue screen. This is especially true with older keyers that are not very selective (the notorious background peeking through a spot of blue on a subject's tie, for example). Modern keyers however are much more accurate and flexible and this is not such a great problem today.

When shooting live actors, a green screen is often preferred. This is due to the fact that any green spill on the subject is much less objectionable than blue. Again however, most modern keyers offer some degree of hue suppression which can greatly reduce this problem. The higher resolution and lower noise of a camera's green channel also favor the use of green screens. Other factors such as the effect of detail correcting circuits in video cameras should also be considered.

Linear and Shadow Keyers

Depending upon the effect required, it is important to choose the appropriate type of chroma keyer. A simple keyer can be used when the subject simply needs to be isolated from the backing and inserted into the composite image. This is often the case when the composite is not meant to be realistic (when the subject will be flown around in the composite using a DVE).

When shooting a live chroma key where the foreground subject contains translucent objects such as glass or smoke, a linear keyer must be used to preserve the illusion. Otherwise, the object simply disappears or worse yet, shows the backing color behind it. A linear chroma keyer reduces the level of the background scene according to the level of the backing color as seen through the translucent object. This makes the composite much more realistic.

Another way to add realism to live chroma keys is to let the foreground subject cast shadows onto the backing (or other keyed out objects in the foreground). Like translucent objects, a linear keyer will recognize the reduced level of the backing color and darken the background video accordingly. This creates a very real illusion of the subject casting shadows into the background scene.

When the chroma key compositing is to be performed later in post-production, keep in mind that most compositing workstations have relatively simple keyers. Fortunately however, there are numerous tricks that can be used to enhance these effects. For example, it may not be possible to pull shadows directly from the foreground video. To get the same result, it may be necessary to create a separate shadow matte, possibly requiring the use of rotoscope. Another trick—although somewhat of a cheat—is to take the shadowless matte and pass it through a DVE to lay it down in the proper perspective; when mixed with the original matte signal, the very effective illusion of a shadow can be obtained.

Once again, the foremost goal when compositing with a chroma key is to reduce any edge effects. This may take a careful balance between all of the keyer and hue suppression controls. Another technique for improving chroma keys in post-production is to add a slight blur to the matte signal. Even better, some devices have the ability to add or subtract one or more pixels around the edge of a matte signal. Such a grow (or shrink) function can be very effective in creating the perfect composite.

Matching Foreground and Background

After edge effects, nothing spoils a chroma key shot more than mismatched foregrounds and backgrounds. The single most important consideration in matching both foreground and background scenes is *lighting*. Obviously the intensity and direction of key lighting in both scenes should be similar. Color temperature and atmospherics should also be matched whenever possible. For example, when the background scene is an outdoor shot on a sunny day, it might be best to shoot the foreground blue screen outdoors as well under similar lighting conditions (pay close attention to shadows).

Other camera parameters such as depth of field may also have subtle effects on the composite image. If the foreground subject is to be placed well in front of the background scene, consider shooting the background with the camera focused onto the area where the foreground subject will later be placed. If this is not possible, the background may be defocused later during compositing to give a similar effect. Once again, pay attention to the details.

Another quality which must be matched between foreground and background is *colorimetry*. Since each scene might be shot with a different camera, lighting setup, film/video setup, etc., it will be necessary to color correct each scene so that they match. Sky colors, skin tones and shadows are usually first to be tweaked. Color correction capabilities vary widely from system to system, but they are very important in creating that seamless composite.

Film Versus Video

Film shoots introduce several other factors to consider. Most importantly, when making the telecine transfer of either background and/or blue screen film to video, a pin-registered gate must be used to eliminate film weave. *Film grain* is another quality which must be matched if the foreground and background scenes are to appear as if they are parts of the same scene.

Motion Considerations

When the effect calls for a moving chroma key, there are even more details to watch out for. Even though a tracking keyer or motion control rig will be used to keep the foreground image in the right place with respect to the background, optical parameters such as perspective and motion blur must also be considered. This is especially true when either the foreground or background scene is shot with miniatures. In this case, camera height and distance to subject should be scaled accordingly.

When an object is moving fast enough so that it would not appear sharply defined in a single frame, there should be some degree of blurring in the direction of travel. Many video cameras have a high enough shutter speed to mask this effect. When called for in the composite, or if the foreground is moved around with a DVE, it may be desirable to add a degree of motion blur electronically to enhance the effect.

Other Effects

Full featured compositing systems will also have the ability to process video clips in various ways. This might include the ability to de-interlace/re-interlace or cine compress/expand a clip. The first function creates two frames from each original frame—one for each field. The latter is used to remove or replace the 3-2 pulldown used in a telecine to convert 24 fps film to 30 fps video. Both of these functions can be useful for rotoscope.

It is also possible to stretch a video clip to make it fill any given time period. Much like an interpolator, the output frames are calculated from the input clip by mixing adjacent frames in proportion to their proximity to the desired time slice. This is often used to add extra speed to a fast scene or to present something in slow motion. There is a limit however to how much a clip can be slowed down. Just as the DVE loses resolution when blowing up an image, a slowed down clip may look strange if the added dissolves between frames become noticeable. If very clean slow motion is required, then the original footage must be shot at a higher than normal frame rate (either high speed film or super slo-mo video).

Motion Tracking

Another useful tool for creating special effects is *motion tracking*. This can be used to composite a new image onto an existing object in a scene that is moving or in a scene where the camera moves. The computer uses a high contrast edge or point to track the movement over a series of frames. This *motion data* can then be used to stabilize the image. This is an effective way to remove film weave from non-pin registered film transfers or to eliminate hand held camera shaking. This process can even be repeated a second or third time to remove any residual motion.

The motion data can also be applied to another image which is to be keyed into the original scene. By tracking the motion of the background scene, the new keyed in foreground element appears to blend into that picture. When done correctly, the tracking motion reinforces the illusion and makes for a great effect; it only takes a small amount of mistracking (wandering) however to give the effect away.

Motion tracking with one point only lets a key follow horizontal and vertical movements. If the object being tracked rotates or changes size, then the keyed in video must be adjusted accordingly. This can be accomplished using *corner pinning* and *four-point tracking*. With corner pinning, a roughly rectangular area is defined on both the tracked background object and the insert video element. The computer then uses its built-in DVE to match the perspective of the insert video by pinning its corner points onto the corresponding points in the background image. If four-point tracking is then applied to independently and simultaneously track all four corner points, the result can be quite spectacular. This technique is often used to track images onto a sign or video screen on the set.

Rotoscoping

Shooting an object or person against a blue screen makes it very easy to generate a matte for the object. This matte can then be used to composite the image over a different background. However, it is not always possible to have the subject shot blue screen. Sometimes, it is necessary to generate a matte for one object within a busy picture. In such a case, present computer tools are usually unable to perform this separation. It then becomes necessary for a human artist to painstakingly create the matte on a frame by frame basis. This is called *rotoscoping*. While mostly used for generating a traveling matte, roto can also be used for other effects such as wire removal and adding hand drawn objects such as sparks, pixie dust or laser beams.

Virtual Sets and Ad Replacement

One technology that appears to hold the most promise for the future of video special effects is that of *virtual sets*. A virtual set is basically a CGI background scene created entirely within a high speed supercomputer. Foreground subjects are then shot blue screen and chroma keyed into the virtual background. The real trick is to then have the CGI background automatically track movements by the foreground camera in real time. Here is where the supercomputer comes in—it must calculate the position and field of view of the real foreground camera and then render a complete frame of the background scene at the proper perspective. And, of course, it must do this at 60 fields per second.

Current computer technology limits the complexity of the background which can be rendered in real time. Thus, early virtual sets were somewhat surreal. But with sufficient computing power, it will be possible to create truly photorealistic imagery (current computers require on the order of minutes—perhaps even hours, for HDTV or film—to create just one such frame).

Creating a virtual set requires two basic components: *camera tracking* and *real time rendering* of a 3D graphic database. Camera tracking—including position, height, pan and tilt angles, lens focal length and even focus—is derived either by position sensors or via pattern recognition. The former method involves the physical attachment of various devices to the camera to relay pertinent data to the render computer. These devices must be calibrated in advance to give the computer a starting frame of reference.

Pattern recognition offers an alternate technique for gathering the camera parameters. With this system, the blue screen is crosshatched with various sized lines. These lines are painted in the same hue, but slightly darker than the regular background. The colors are close enough to be completely keyed out together, but different enough that they can be detected. The camera parameters can then be extracted by analyzing the camera's video signal. By detecting the presence of and the relative relationship between, various lines of the crosshatch pattern, the computer can determine all of the needed parameters. Even the fuzziness of the lines can be used to determine the relative focus needed to render the background correctly.

The camera tracking technology of virtual sets was first used for advertisement replacement during sporting events. It is now quite possible that a banner or logo seen in the stands (or even on the field, court, rink, etc.) of a televised sporting event might not actually be visible to a spectator at the same event. And pattern recognition can also be used to create other interesting effects (tracking the motion of a single player or other object in a live sports production).

The use of virtual sets is currently limited to special news shows, some sci-fi and a few children's programs. As the technology improves however, it may see wider use. If for no other reason, the potential for reducing set construction and storage costs make it fiscally appealing. Several networks have given considerable thought towards its use in producing soap operas, for example. However, one major stumbling block still remains—it requires a different kind of acting to play a scene inside a virtual set. Gone are the traditional walls—and in some cases even the props—that an actor must interact with.

Some believe that the next great visual effect will be the use of *virtual actors* (dead actors brought back to life or new, totally computer generated actors). Currently, live actors wearing body suits with motion sensors can be used to animate CGI characters. Higher resolution however is needed to capture slight nuances such as facial expressions. In the future it may be possible to eliminate the actor altogether and have a realistic character completely controlled by computer.

VIDEO TO COMPUTER INTERFACES

The power and low cost of personal computers has not gone unnoticed by video professionals. Some of the most exciting and lucrative fields of software development are in the area of video creation and processing. Many tasks that are traditionally performed on dedicated video production equipment or 3D supercomputers can now be executed on low cost PCs.[2] Of course, it may take much longer to perform the same task on a PC, but often this is well worth the tradeoff of much lower equipment costs. And as PCs get faster and more powerful every year, the gap between them and dedicated equipment continues to narrow. With the addition of one or more specialized plug-in boards (or an external video processing unit), this gap can even be eliminated; indeed, many new video production tools are now based on, or *driven by*, PCs. With the advent of digital television (DTV), this approach will undoubtedly be used even more.

[2] The use of the term personal computer PC here is not platform specific.

Video Format Conversions

To process video in a computer requires some manipulation of the video signal. Analog composite or component video must obviously be digitized; composite signals must also be decoded into RGB or YUV components. Even component digital video signals may require YC_bC_r to RGB color space conversion and/or pixel aspect ratio correction. For compressed video systems, conversion to an M-JPEG, MPEG-2, or similar bitstream must also be accomplished. And these processes will likely need to be reversed when bringing the video back out of the computer realm.

To understand all of the conversions necessary, it is helpful to examine each of the standards currently in use. Broadcast engineers should be familiar with the NTSC composite, RGB analog and perhaps the ITU-601 digital video standards.[3] Inside computers, video is most likely to be represented by 24 bit RGB values. An optional 8 bit alpha channel is sometimes added to ease keying (besides, computers handle 32 bit values more efficiently than 24 bit). (The NTSC composite video format is described in detail in Chapter 1.9, *NTSC Standard*). The RGB analog format utilizes three separate monochrome video circuits to convey the complete video signal.[4] Sync and setup are optional on each signal and obviously there is no color burst. If sync is not present on any of the RGB signals, it must be carried as a fourth signal and the format is referred to as RGBS. Since this adds to the complexity of the overall system, most equipment is designed to add/extract sync from the green channel, otherwise known as *sync on green*. Sync may also be added to the red and blue channels to prevent discrepancies in the way the three channels are processed; however, this sync is never used.

Component digital (ITU-601) is based upon YC_bC_r component signals.[5] These are derived from the standard, gamma corrected R′G′B′ signals according to the following equations:

$$Y = 0.257R' + 0.504G' + 0.098B' + 16$$
$$C_b = -0.148R' - 0.291G' + 0.439B' + 128$$
$$C_r = 0.439R' - 0.368G' - 0.071B' + 128$$

By definition, the Y signal has a range of 16-235 and the C_b/C_r signals have a range of 16-240 (assuming 8 bit resolution). This leaves some digital headroom for over- and under-shoots of the video signal. Even so, much of the YC_bC_r color space is outside of the standard RGB gamut, as shown in Figure 5.9-4. This is a result of restricting the RGB values to a range of 0 to 255.

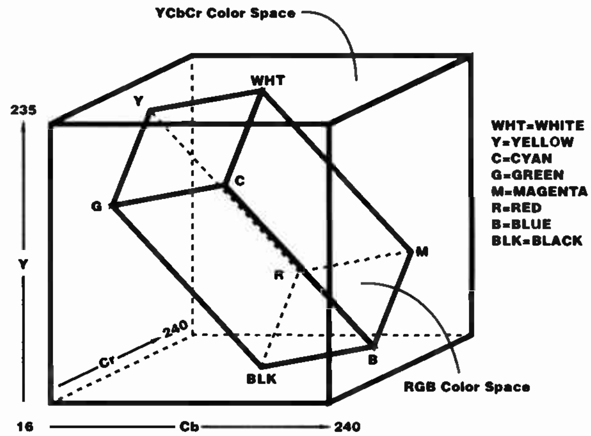


Figure 5.9-4. Comparison between the YC_bC_r and RGB color spaces. Note that about one-half of the YC_bC_r values are outside of the RGB gamut.

To make matters worse, some RGB combinations translate into illegal colors when encoded to NTSC composite (mostly due to excessive chroma levels). For this reason, it is imperative to check levels on all signals brought in from YC_bC_r space (from a digital paint system). Any objects created on the computer should be drawn using an NTSC legal palette. Compositing programs should have an NTSC filter function to ensure that their rendered output is within spec.

Picture Resolution and Pixel Aspect Ratio Conversions

Picture resolution (both horizontal and vertical) and pixel aspect ratio must also be considered when interfacing video signals with a computer. This is most likely to be an issue when dealing with ITU-601 signals. The ITU-601 standard specifies a horizontal resolution of 720 pixels by 486 lines.[6] Given the 4 × 3 (or 1.33) aspect ratio of NTSC, this effectively makes each pixel rectangular. Computer graphics adapters however, almost exclusively use square pixels. Since the number of lines already matches the number of active lines of video in 525 NTSC, it remains a logical choice. Therefore, to give a true 1.33 aspect ratio using square pixels, there would have to be 648 pixels per line. To convert between 720 × 486 and 648 × 486 requires that the pixels be squeezed horizontally by 10% (multiplied by 0.9).[7] The 648 × 486 resolution

[3] Formerly CCIR 601, now completely specified as ITU-R BT. 601-4 (1994) *Encoding Parameters of Digital Television for Studios*.

[4] RGB may alternately be denoted as GBR.

[5] YC_bC_r is the correct terminology for this luminance (Y), R-Y (C_r) and B-Y (C_b) component format. This is sometimes incorrectly referred to as YUV (which also is related to the unscaled and offset signals upon which YC_bC_r are derived).

[6] Technically the standard allows for 487 active lines in NTSC. When dividing the 525 lines per frame into two fields, the standard does not allow half-lines. Thus an extra line is added to field one. Whether this line is used for active video is somewhat debatable, but the majority of ITU-601 devices appear to specify an image size of 720 × 486.

[7] Again, there remains some debate over the exact pixel aspect ratio. While ultimately this remains a function of the specific display device, some calculations prefer to eliminate a few horizontal pixels which get blanked, or are used for the transitions into and out of blanking. It is common to see the number of "active" pixels per line set at 711, which translates into a pixel aspect ratio of 0.911.

is often further scaled down to 640 × 480. Although this is somewhat easier for the computer to handle, it does not allow enough active vertical lines to create a legal (as defined by the FCC) signal. Despite this however, many 640 × 480 images are still being broadcast today.

When the final result is to be laid off to film, the computer must render a higher resolution image. Typically, film images will be handled at *4K resolution* which implies 4096 pixels per line. Since film aspect ratios vary considerably, this could represent a frame size between 4096 × 1870 (70 mm @ 2.19:1) to 4096 × 2990 (35 mm @ 1.37:1). At 24 bits per pixel, that translates into a file size of 22–35 MB per frame. This compares to approximately 1 MB for a single frame of 720 × 486 video.

Interfacing Video with Computer Workstations

For offline quality editing, composite video will probably suffice. Even online editing is possible thanks to faster computers and hard drives which allow for lower compression ratios. Special effects and compositing however, are usually done in an uncompressed, component video domain. For most television work, this will be based on the ITU-601 Standard.

Images created solely within the computer however can be rendered to any resolution. Higher resolutions of course, will require corresponding increases in rendering time. For this reason, long video clips (or short, film resolution scenes) can require rendering overnight. While this may make good use of computing time, it also carries the danger that something will go wrong which halts the process. Or, after seeing the final render, there may be a problem that requires making a change. In either case, another long render may be required. For those working under tight deadlines, this can be disastrous.

Composite video work can be digitized and laid back out in real time. Film resolution jobs, on the other hand, will usually be laid off from the computer to some form of digital tape (Exabyte™), or possibly transferred directly over a network to the digital film recorder. Most video work however, is done by transferring component video in and out of the computer. This can be accomplished in one of three ways: using *analog component video, digital component video* or via *data transfer.*

Analog component video usually requires an add-on video card that feeds RGB (or YC_rC_b, Betacam™, etc.) to the computer. This card performs sampling, digitizing, and possibly color space conversion on the input side. For output, it must perform D/A conversion and filtering. Make sure that the card is set up correctly for the type of component signals to be used (whether setup is present or not on RGB signals). Also be careful when feeding the analog output from a computer into other ITU-601 devices. Images created on the computer often have risetimes exceeding the standard NTSC bandwidth. When these signals are fed into a digital device, they can cause ringing on sharp edges. Some computers also have built-in S-video ports; while certainly better than standard composite, they are still not as clean as full component video.

Feeding component digital to the computer requires a very high end video card. Serial digital video (SDI, per ANSI/SMPTE 259M) most likely will be color space converted between YC_bC_r and RGB. The levels may also be scaled to match the computer ($Y=16$ black would scale to $Y=0$, or $R=G=B=0$; $Y=235$ white becomes $Y=R=G=B=255$). Color bandwidth is expanded from 4:2:2 to the computer's equivalent of 4:4:4 sampling. Although these operations are mostly transparent, beware of illegal colors and optional filters that can cause slight variations in color. This could make the processed frames stand out if they are edited back into the original clip. The best advice is to work in YC_bC_r space whenever possible.

As with most UNIX-based 3D workstations, video frames can also be passed as data files. These files can either be written to a standard data tape format or transferred over a network to another digital device. Quite often, the other device is a DDR. In this case, the video is rendered and transferred to the DDR one frame at a time. When this process is complete, the entire clip can then be played back in real time on the DDR.

Video Editing/Compositing Programs

Once video has been digitized and stored within the computer, it is up to a software program to perform whatever manipulation is required. This could be as simple as cuts-only editing or as complicated as full multi-layer compositing with DVE moves, morphs, motion tracking, etc. Most programs offer an intuitive *user interface*, often representing clips as reels of video frames (like filmstrips) and using menus and buttons to operate on these clips. One of the main advantages of general purpose PCs is their ability to run several different programs on the same hardware. Therefore it is possible to capture a video clip with one program, send the video to a standalone morphing program and finally composite the morph using yet another piece of software.

Video editing programs are most likely to use compression techniques to store the lengthy video clips used in long form production. High end systems will use the least compression (approximately 2:1) for online work. With additional hardware, it is possible to process two video streams simultaneously, allowing dissolves and simple effects to be created in real time. These programs also handle multiple channels of digital audio at the same time.

For effects work, 2D and 3D animation programs are often used in conjunction with compositing software. These programs tend to work with higher quality, uncompressed video. As such, they do not operate in real time. Although this type of work generally deals with short video clips, even a few seconds of complicated compositing can take an hour or more to render.

KEY TERMS

CGI. Computer Generated Images. Still pictures or complete motion video created solely within a computer (i.e. without a camera).

Compositing. The layering of multiple video elements to create a montage of motion video.

DVE. Digital Video Effects creator. May be a standalone unit, combined with a switcher or part of an integrated video workstation.

Field dominance. The specification of which field (one or two) is to denote the start of a new frame. Most facilities adopt the standard of field one dominance. Incorrect field dominance can result in juddery motion and flicker frames at edit points.

Fill. The foreground video signal which is keyed over a background using a matte.

Hicon. High contrast matte. A monochrome key signal ranging from black (making the fill totally transparent) to white (making the fill totally opaque).

In-betweening. Creating those frames in between keyframes of an effect.

Interpolation. Calculating the value of a pixel based upon its surrounding pixels. Also used to describe the creation of frames when altering the speed of a clip.

ITU-601. Formerly CCIR-601, the standard for component digital video. Specifies a resolution of 720 × 486 (NTSC), 4:2:2 sampling (chrominance sampling at one-half that of luminance) and 8- or 10-bit signaling. Also the format used internally by many digital video devices, most notably the D1 VTR (See endnote 3 for full title to ITU-601).

Matte. Also referred to as a key, is the signal used to determine which parts of the foreground and background images appear in the final composite. (Also used to describe to the keying process itself.)

Keyframe. A specific frame of an effect where certain parameters are specified.

Rotoscoping. The frame by frame hand drawing of a video clip. The drawing may be used to create a moving matte or to add special hand drawn effects to a clip.

BIBLIOGRAPHY

A60 Interface Handbook, Abekas Video Systems, 1992.

"GREEN OR BLUE—Selecting a Backing Color for an Ultimatte Composite," *Ultimatte Technical Bulletin No. 2*, Ultimatte, Chatsworth, CA.

ITU-R Recommendation BT.601-1994, *Encoding Parameters of Digital Television for Studios*, ITU.

Keith, Jack., *Video Demystified, Second Edition*, HighText Books, Solana Beach, CA, 1997.

"Matching Foreground & Background in a Composite," *Ultimatte Technical Bulletin No. 7*, Ultimatte, Chatsworth, CA.

Mazur, Jeff., "Compositing and Effects," *Broadcast Engineering*, Intertec Publishing, Overland Park, KS, pp. 48–54, November, 1996.

Pank, Bob., *The Digital Fact Book, Edition 8*, Quantel, 1996.

"Shooting Film for Ultimatte," *Ultimatte Technical Bulletin No. 3*, Ultimatte, Chatsworth, CA.

SMPTE 125M-1995, *Television—Component Video Signal 4:2:2—Bit-Parallel Digital Interface*, SMPTE.

SMPTE 259M-1993, *Television—10-Bit 4:2:2 Component and 4fsc NTSC Composite Digital Signals*—Serial Digital Interface," SMPTE.

"Ultimatte vs. Chroma Key," *Ultimatte Technical Bulletin No. 5*, Ultimatte, Chatsworth, CA.

5.10
STATION AUTOMATION AND NETWORKING

JOHN D. WEIGAND
KSWB-TV SAN DIEGO, CA

INTRODUCTION

When looking at an overview of automation for television stations, the network computer system and the various departments within the infrastructure of the station must be considered. Each department must be carefully reviewed in order to create a well running, efficient station-wide system. Engineering, traffic, programming, accounting, promotion, sales and news each have unique and specific needs that must be met, and the automation process must integrate computer hardware and software used within the various departments in order to create a station wide computer system network (see Figure 5.10-1).

Network Architectures

The basic computer system within a station requires a local area network (LAN). The LAN most frequently used today is the 10Base-T system. The station with foresight will use hardware allowing for future expansion to a 100Base-T system. This is because as more departments are added to the network and more data is required through the LAN, the LAN slows down. The faster the system is, the greater is the amount of information that can be sent upon it. The 10Base-T system allows information to travel at 10 Mbps, while the 100Base-T system moves at 100 Mbps. If the station can afford to build a 100Base-T system to begin with, it will be better off in the long run. However, building a 100Base-T system is far more expensive than building a 10 base-T system and is, in most instances, not a cost effective alternative with which to begin.

A compromise between these two alternatives is to install hubs and group lower volume users on the same hubs, while placing higher volume users on faster hubs. The hubs can then be connected though a multiport switch. This will save money on a short term basis and still keep data moving through the LAN system quickly.

How can a station plan for future expansion to a 100Base-T system? Any network interface cards installed should be a 10/100Base-T interface that will work at either a 10Base-T speed or a 100Base-T speed. All wiring used should be category five wire that will support operation of a 100Base-T system. In this manner, the foundation will be laid for moving to a 100Base-T system in the future.

When building a new computer network within a station, or when upgrading an existing network, planning should include switched networks in order to reduce system latency. Multicast and automated push technology, and accompanying increased traffic, will bog down the network. Figure 5.10-2 shows a typical stacked hub network where all ports share on a 10

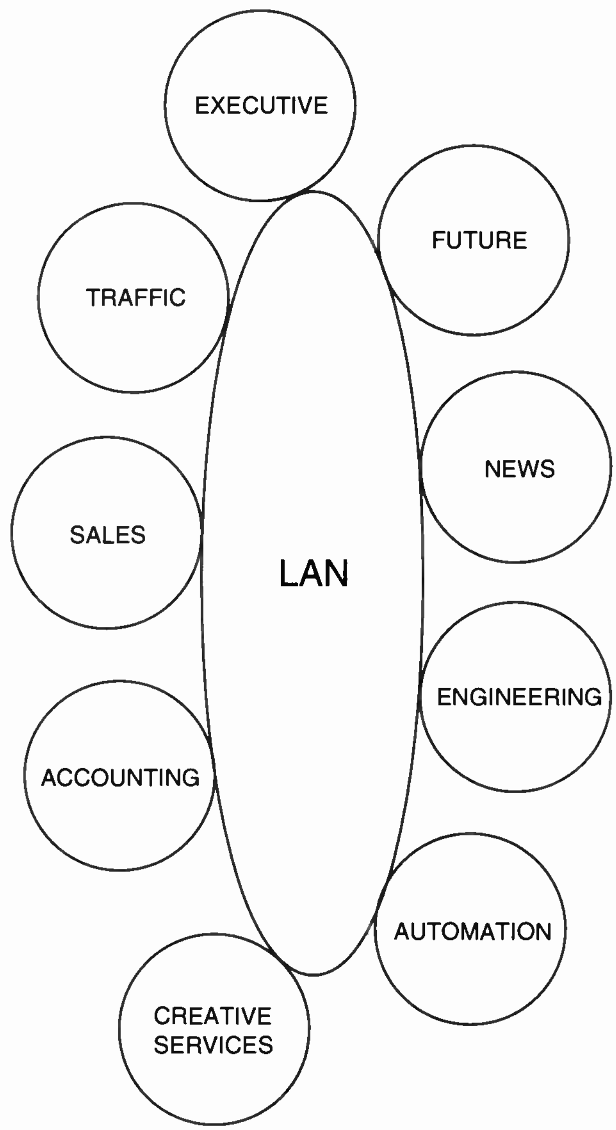


Figure 5.10-1. The elements involved in a television station automation system.

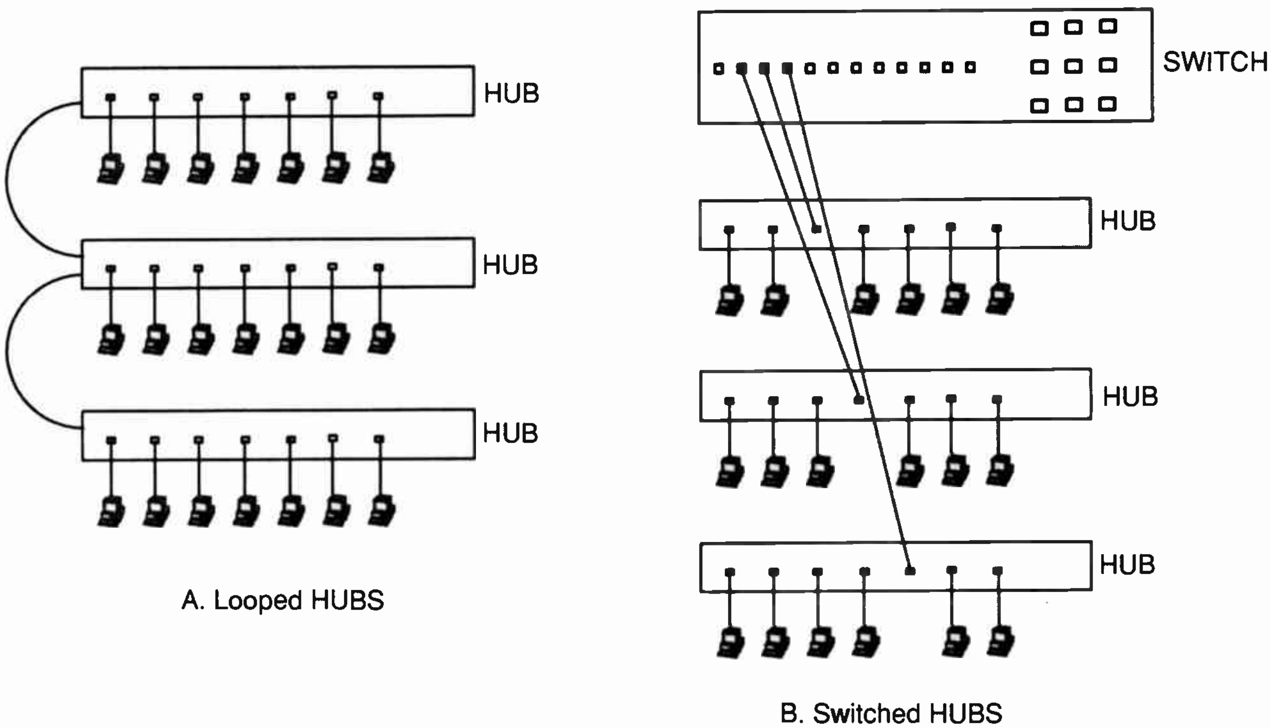


Figure 5.10-2. Network hub configurations: (*a*) looped hub, (*b*) switched hub.

Base-T system (10 Mbps). Adding a switch where each hub sits at a separate switched port reduces *latency* of the system. Latency is the slowing down of the network and it must be avoided whenever possible and practical. The intelligent choice for a computer system within a television station is, therefore, the switched system that avoids latency within the network.

In the past, servers have been connected through a hub to the backbone, or LAN, where traffic is divided by network components such as bridges, routers and switches. These components create *firewalls* between sub LANs, filtering out unwanted traffic. However, from the hub to the backbone, these servers share the 10 Mbps connection up to the backbone. This results in a number of servers working through the same restricting element, as illustrated in Figure 5.10-3. In order to increase the bandwidth, each server should be connected to a switch having multiple switched connections to the backbone, shown in Figure 5.10-4. Each server will then have 10 Mbps *dedicated bandwidth* (each server has a separate a connection to the backbone). If there are three servers, there should be three connections to the backbone, and so on. This increases speed of the network as well as its overall efficiency.

In the network architecture, there should be a grouping of users by *commonality*. This means that traffic, accounting, general office, programming, sales, promotion and engineering should ideally be placed on a separate hub, and from there each group or hub should be connected through a switched port. Such an approach will reduce the latency of the network.

The IEEE Standards P802.1P and P802.1Q prioritize the class of service and quality of service within a network by addressing the issue of separate cueing of time critical frames in bridges, in order to cut down

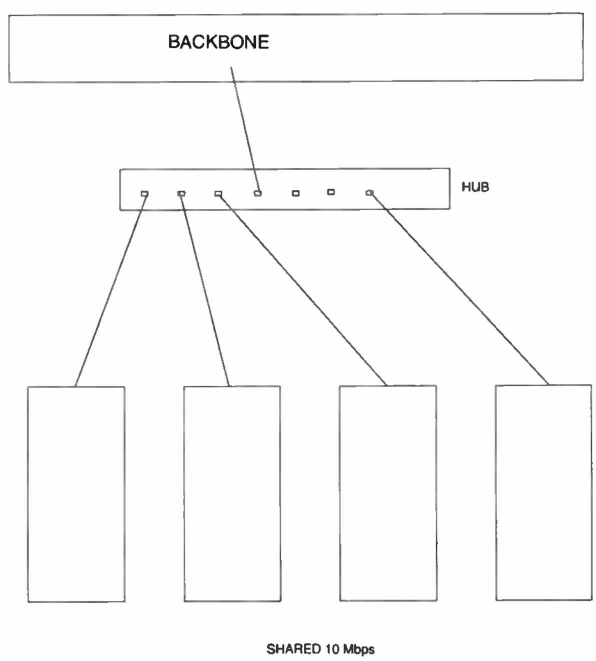


Figure 5.10-3. Network choke point caused by a single connection from the server hub to the backbone.

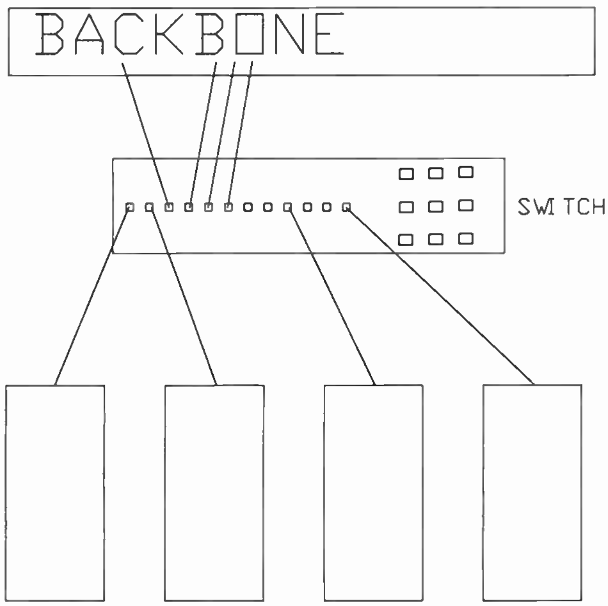


Figure 5.10-4. Multiple switched connections to the backbone, which provides each server with 10 Mbps dedicated bandwidth.

on the latency imposed by the handling of large data blocks. This will allow dynamic configuration and distribution mechanisms to work in such a manner that *general attribution recognition protocol* (GARP) can be used to reconfigure video LANs (VLANs) and propagate information databases.

Push technology allows each user to receive automated transmissions that are broadcast to the entire network. A problem often occurs with push technology and the standard unswitched 10baseT network because they can cause the network to bog down as it sends the information to each user. This problem is avoided simply by moving to a switched network and utilizing the attributes of P802.1P and P802.1Q.

With everything that is being sent over today's networks—video conferencing, stock quotations, Internet research information and the like-television station planners must carefully consider growth options needed for the future of the facility. They may do this by researching and keeping abreast of present and future developments in this important area of technology.

TELEVISION EQUIPMENT AUTOMATION

The standards for television equipment automation include Esbus, RS232 and RS422. A great deal of the television equipment available—VTRs, servers, satellite receivers, satellite controllers, switchers, routers, character generators and related hardware—adhere to these standards. When manufacturers design equipment, they choose which standards they will use. Most of the equipment available is set for RS422. However, it should be noted that RS232 equipment can be converted to the RS422 Standard and vice versa. Esbus is a higher form of remote control and is the SMPTE preferred protocol for RS232 and RS422 bidirectional inputs and outputs.

Protocols are tables of commands and status reports that videotape machine manufacturers develop for their products so that outside devices, such as remote control units or editing units, can tell what is happening. Protocols complement the remote control capabilities of the machine and tell the remote control system the status of the machine. In the past, some manufacturers of tape machines were reluctant to share protocols with competitive manufacturers of units used in conjunction with their tape machines, which made planning difficult. However, that is not the case today. Automation tables are commonly published on computer software within the automation system. The information is readily available when performing the initial set up of the equipment-run software, making planning easier.

Automation within Departments

Automation in the engineering department means controlling satellites, VTRs, servers, switchers, routers, character generators and archiving of commercial and promotional material.

Satellites can be automated by interfacing the antenna controller and the satellite receiver to the automation system. This enables the user to program timed rotation of the antenna for a specific satellite and polarity, either vertical or horizontal. The satellite receiver also is controlled to select the transponder and audio subcarrier channels. The video and audio may be directed to a recording device through control of the router. For example, if the user is utilizing a tape machine, the automation system would confirm there is a tape machine and that a tape is ready before beginning to record.

VTRs can be controlled through automation or manual control. In an automated system time code is read off of the videotape by the VTR. The automation program uses the time code read from the tape to determine exactly when to switch the VTRs output to the program channel. Clearly, in this situation, the automation system needs to know or at least needs to calculate the time code reading at the exact point on the tape where the program begins.

Automation control servers through RS422 ports, and read the timecode on the servers to find specific pieces of video (or audio).

Automated master control switchers receive the information they need to control cross-point switch closures at scheduled times from the traffic/automation system via the LAN or *sneaker net* (manually transferred data). The same can be true for all automated routing switchers.

Character generators are also controlled though RS422 control ports and in this manner they can be used to automate control character output consisting of station identification, transparent branding (bugs) and other information.

Software written for automation can include data that enables a copy of commercial or promotional material to be stored in an archive, generally a digital storage tape.

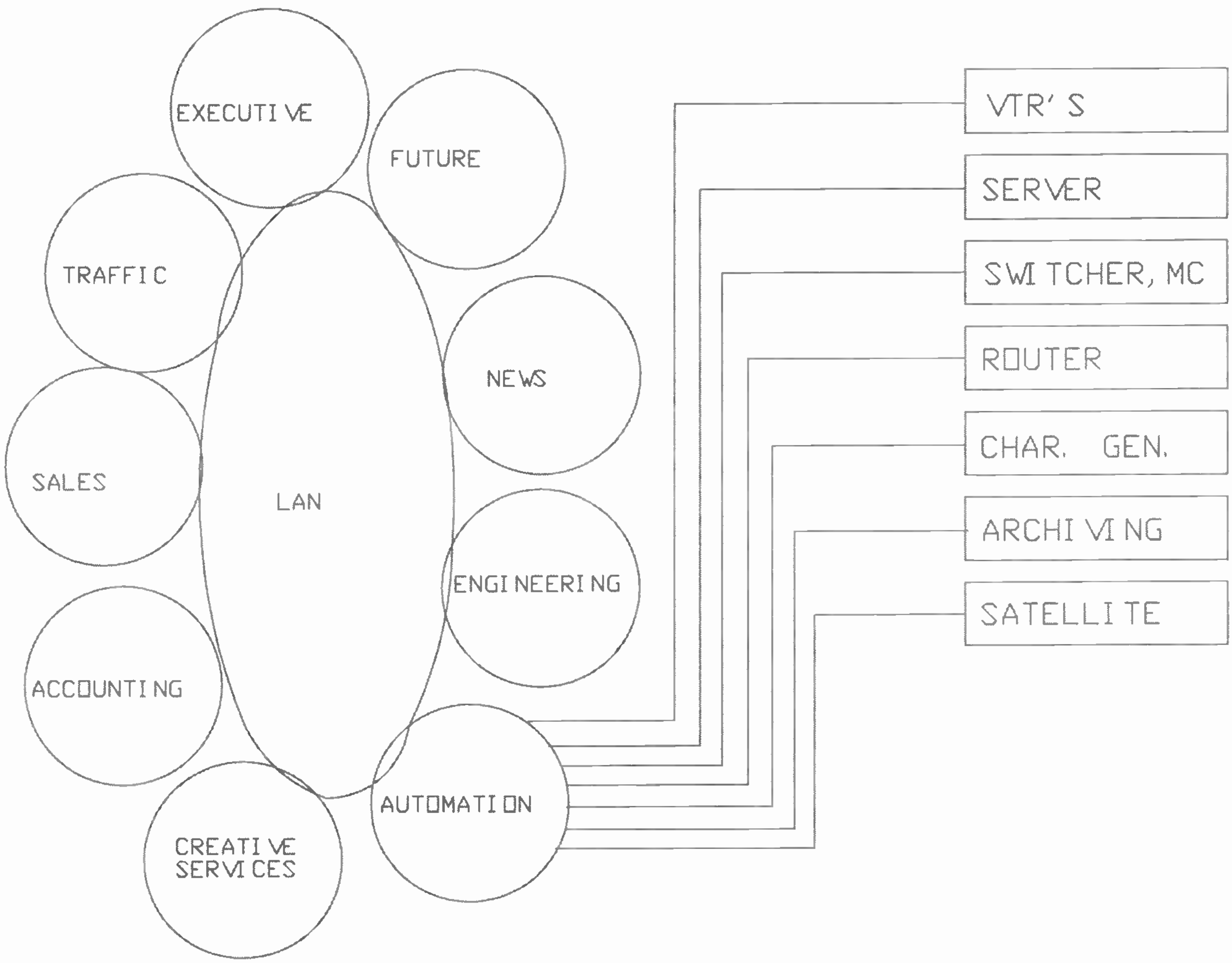


Figure 5.10-5. Integration of a television station automation system with various facility functions, equipment and departments.

All of the broadcast equipment discussed in this section is directly controlled through a single automation computer, as illustrated in Figure 5.10-5.

The traffic department can use contracts from sales for commercial playback scheduling. It can schedule program content received from the programming department and create an entire program for the broadcast day. The accounting department can reconcile the program log. The promotions department can create promotional materials to run on air. Clients can be invoiced from the program log for commercials that have actually run on-air. This multitasking can be done through a building-wide network. All departments are connected on the same LAN, but through different switch ports and servers. The computer use is then interfaced with the broadcast equipment systems. While it would be useful to do programming of broadcast equipment directly though the LAN, at this time some manufacturers recommend the sneaker net. Using the recommended sneaker net means the user hand carries the daily log down to the engineering department on a floppy disk. This disk is installed in the dedicated equipment computer, rather than launching the information through the LAN. The reason for this approach is usually the fear that the program log might become corrupted at some point in the LAN-to-broadcast equipment chain. Log corrections on the day of the log are done from a traffic workstation sent through the LAN to the automation system, which converts the file, makes the corrections and notifies the master control operator of the change on the screen.

In order for traffic functions to work efficiently with the automation system, there must be a 24 hour log. The traffic log should be able to use six digit numbers. The lowest order digits (in the ones and tens places) should be reserved for the cut number of each segment. This way, it is easier to keep track of the client's multiset cuts. Also, for promotional material, there can be a two week rotator based on week "A" and week "B." This will give a seven day time frame to input the "B" week while the "A" week is still running. This facilitates technical operations interfacing with the traffic and promotions departments, permitting easy updating of the various schedules.

The creative services departments can move nonlinear and computer graphics through the LAN to other departments, or from work station to work station for nonreal-time processes over the 10Base-T Ethernet LAN or 100Base-T hub on a switched port of the LAN.

The sales department can schedule air time for a sale and easily access the LAN to see where time is available. At sales sites, the marketing representatives can call into the station network, access traffic, and immediately see what time is available for sale. This facilitates immediate sale of the time slot. At the same time, the market representative of the station can communicate with traffic regarding the sale and electronically enter the contract into the accounting traffic system, which all resides on one common LAN.

Networking and Automating the Newsroom

The network system for the newsroom should be both *modular* and *scaleable*. At this writing, newsroom automation and network systems are in a major transition. An example layout of a typical newsroom of the near future can be seen in Figure 5.10-6. In this newsroom, news stories are gathered though satellite and recorded on a server in real-time. Stories are gathered on a field unit (recorded on a camera VTR storing information digitally on the camera hard drive), which can be played back at four times real-time and then fed to the studio server, either physically or by microwave relay. An abbreviated version of the stored stories is fed into a clip server. The clip server feeds the various work stations for the writer, producer or editor for writing and editing the news to be aired. At any one of the work stations, both writing and editing may be done simultaneously. This means the camera person can bring in the camera and edit and write the story alone, or let a writer/producer edit the story as it is written. Thus, the jobs within the newsroom of the future are likely to be redefined. Instant editing and writing of a breaking story as it is fed by satellite or microwave through the server to one or more of the work stations is one reasonable scenario. At any one of the work stations, a newsperson can write the story for the TelePrompTer, indicate and signal the character generator to type keyed characters on air and make an edit decision list. Video can be fed from the server out to a nonlinear edit suite, where A-B edits can be made. Presently, in most newsrooms, video has to be played from videotape machine A and videotape machine B and recorded on videotape machine C in an A-B edit room. With new technology, the process can all be done on one machine, a nonlinear editor.

With camera robotic capabilities in the news studio, the individual writing the story to be read on air is further able to direct camera shots required by first inputting the shots and chromakey positions into the robotics computer and then assigning shot numbers. From this point, the robotics computer can be linked to the newsroom computer. When a story is written at the newsroom computer, the robotic system shots can be programmed by entering the appropriate shot number. The two computers are linked on a LAN, bridged or switched through one of the network hubs.

Stories can be sent from the server to air, and simultaneously archived on digital tape as data within their respective archive machines. The servers and the archive machine are connected by a Fibre Channel, as illustrated in Figure 5.10-6. From this point, scripts, graphics, video and audio information can be sent to a web publisher. This data will be converted into *hypertext markup language* (HTML) and sent to the station's web page.

Management Considerations

Automation can reduce the cost of operating broadcast station facilities, making more effective use of personnel and time. Operating costs can be reduced, showing a healthier bottom line. Instead of each department working solo within the broadcast station, the various departments will begin working together, forming a new cohesive bond as they share equipment resources. The facility investment will be protected by keeping up with the latest technologies available. Income relative to expenditures will be maximized.

The engineering department must begin the automation process with careful planning and evaluation of system needs. Planning needs to include the physical elements required to automate the station, the workflow procedures currently in place and the work patterns of the personnel involved. A traffic system, for example, must be developed with a synergy to engineering, programming and marketing.

While you may be building a new plant and making it digital, keep in mind the current and future requirements of DTV, bring a whole new set of standards and will pose an entire new set of rules to be considered.

DIGITAL VIDEO INTERFACE AND NETWORKING[1]

Parallel connection of digital video equipment is practical only for relatively small installations. There is, then, a clear need to transmit data over a single coaxial or fiber line.[1] To reliably move large amounts of data from one location to another, it is necessary to modify the serial signal prior to transmission to ensure that there are sufficient edges (data transitions) for reliable clock recovery, to minimize the low frequency content of the transmitted signal, and to spread the transmitted energy spectrum so that radio frequency emission problems are minimized.

In the early 1980s, a serial interface for ITU-R Rec. 601 signals was recommended by the EBU. This interface used 8/9 block coding and resulted in a bit rate of 243 Mbps. The interface did not support 10 bit

[1] This section was contributed by Jerry C. Whitaker, editor-in-chief.

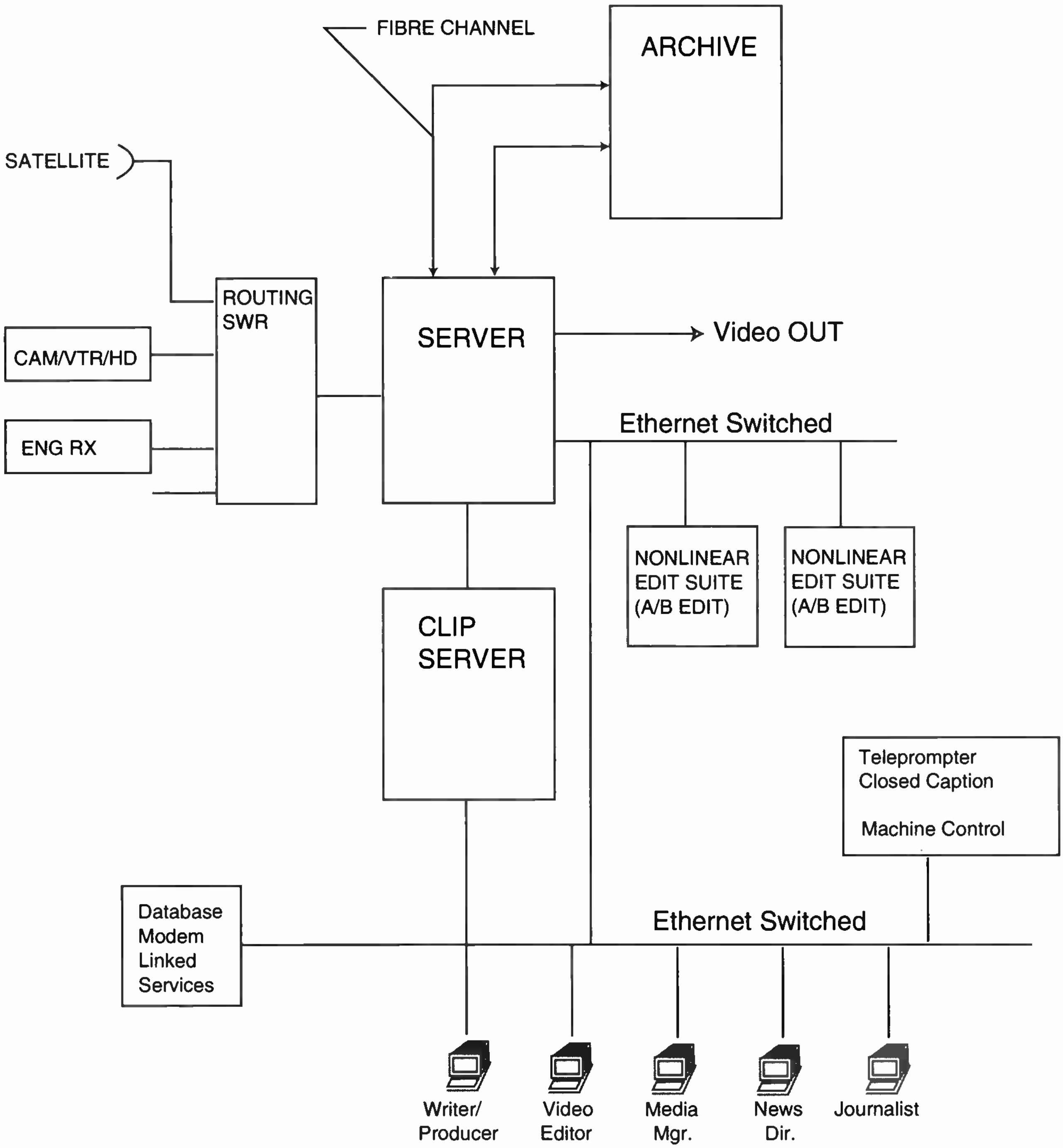


Figure 5.10-6. A full featured newsroom server system.

precision signals, and there were some difficulties in producing reliable, cost effective integrated circuits for the protocol. The block coding-based interface was abandoned and replaced by an interface with a channel coding scheme that utilized scrambling and conversion to *non return to zero inverted (NRZI)*. This serial interface was standardized as SMPTE 259M and EBU Tech. 3267, and is defined for both component and composite conventional video signals, including embedded digital audio.

From this beginning, a number of complementary techniques have been developed to efficiently move digital video information from one location to another. Some of the systems—such as SMPTE 259M—are intended for intra-facility applications, while others—such as *asynchronous transfer mode* (ATM)—are intended for inter-facility applications. The challenge today lies in utilizing and then interfacing the various systems.

Serial Digital Interface

Conceptually, the serial digital interface (SDI) is much like a carrier system for studio applications. Baseband audio and video signals are digitized and combined on the serial digital carrier. (SDI is not strictly a carrier system in that it is a baseband digital signal, not a signal modulated on a carrier wave.) The bit rate (carrier frequency) is determined by the clock rate of the digital data which is 270 Mbps for the ITU-R Rec. 601 component digital. The widescreen (16 × 9) component system defined in SMPTE 267 will produce a bit rate of 360 Mbps. This serial interface may be used with normal video coaxial cable or fiber optic cable, with the appropriate interface adapters.

Following serialization of the video information, the data stream is scrambled by a mathematical algorithm and encoded. At the receiver, an inverse algorithm is used in the deserializer to recover the data. In the serial digital transmission system, the clock is contained in the data, as opposed to the parallel system where there is a separate clock line. By scrambling the data, an abundance of transitions is assured, which is required for reliable clock recovery.

Figure 5.10-7 shows the SDI bitstream for 270 Mbps and 360 Mbps operation. The EAV and SAV elements of the bitstream are reserved word sequences that indicate the start and end of a video line, respectively. For the 270 Mbps case, each line contains 1440 10 bit 4:2:2 video samples. The horizontal interval (HANC) contains ancillary data, error detection and control, embedded audio and other information.

The *serial digital transport interface* (SDTI) is a SMPTE Standard that uses the current SMPTE-259M serial interface link as a carrier for other types of data. It is seen as a valuable component in the distribution of both SDTV and HDTV signals within a television plant. SDTI allows reuse of conventional SDI equipment and wiring to carry compressed video signals. The resulting data stream looks like serial video and, therefore, works with 270 or 360 Mbps SDI. SDTI can, in fact, be used for any compressed video format. Furthermore, it allows faster than real-time data transfer between devices in SDI.

Referring to Figure 5.10-7, in the SDTI case, the active line is the *user data payload.* This payload carries packetized compressed video samples, which are subdivided into blocks of information. In compressed video transport applications, the blocks are fixed in size and carry MPEG-2 transport packets. Header blocks in the SDTI stream contain addressing and other information required for processing of the data blocks. Table 5.10-1 lists the payload capacity of SDTI as a function of format.

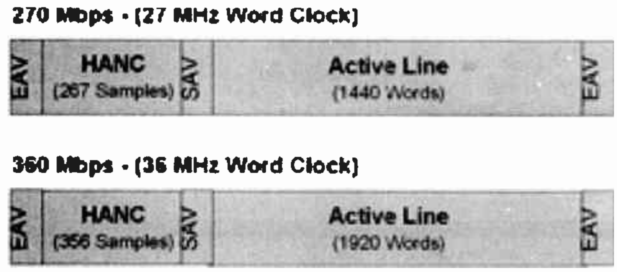


Figure 5.10-7. The basic SDI bitstream. *(Courtesy of Leitch.)*

High-Definition Serial Digital Interface

Based upon the experience gained from the application of digital technology to the 525/625 broadcast and production environments, it is evident that wide acceptance of new professional video equipment takes place only after an efficient means of interconnection is available.[2] For studios today, the means of interconnection is typically SDI. In an effort to address the facility infrastructure requirements of HDTV, SMPTE and Broadcasting Technology Association (BTA) (Japan) developed a standard for digital serial transmission of studio HDTV signals.

The overall transmission rate for transporting a digital studio HDTV signal (1125-line, 2:1 interlace, with 10 bit component sampling) is approximately 1.5 Gbps. The active payload is on the order of 1.2 Gbps (for 1035/1080 active lines). The transmission of video signals at these bit rates represents a far more difficult technical challenge than serial distribution at 270 Mbps used for conventional television signals.

The introduction of the SDI for conventional video in 1986 (SMPTE 259M) was well received by the television industry and has become the backbone of digital audio/video networking for broadcast and post production installations around the world. SDI is ideally suited to the task of transporting uncompressed component/composite digital video and multichannel audio signals over a single coaxial cable. To emulate the same level of operational usability and system integration of conventional television equipment in the HDTV world, the implementation of a *high-definition serial digital interface* (HD-SDI) system—based on an extension of SMPTE 259M—was essential.

Work on the HD-SDI system began in 1992 under the auspices of SMPTE and BTA. The end result of these efforts was the BTA document BTA S-004 (May 1995), followed closely by SMPTE 292M (July 1995), both with similar technical content.

Table 5.10-1
Payload Capacity.

Format	Serial Data Rate	Payload Per Line	Approximate Maximum Transport Payload
SDI, SMPTE-259, Level C	270 Mbps	1440 10 bit words	225 Mbps
SDI, SMPTE-259, Level D	360 Mbps	1920 10 bit words	300 Mbps

The source formats of SMPTE 292M adhere to those signal characteristics specified in SMPTE 260M and 274M. In particular, the field frequencies of 59.94 Hz and 60.00 Hz, and active line numbers of 1035/1080 are used by HD-SDI. Table 5.10-2 lists the basic parameters of the input source formats.

Network Contribution Options for DTV

The form in which material is delivered to broadcast affiliates is a point of some interest and concern as stations plan their DTV plant. The simplest approach has the station taking a broadcast-ready, 19.4 Mbps ATSC data stream from the network. This feed, however, is basically acceptable for pass-through functions only because of the compression already taken on the signal. For applications requiring some form of manipulation of the incoming signal, a higher bitstream rate is required. Options include the following:

- 45 Mbps over satellite or telco fiber links
- 68 Mbps from satellite systems using the latest in modem technology

There are also several potential negative implications of a higher bitstream, including:

- It requires one satellite transponder per signal
- Requires a contribution-grade encoder/decoder
- If some affiliates simply pass-through, they will still require a 19.4 Mbps bitstream
- Concatenation artifacts may arise from different compression systems.

Mezzanine Level Distribution

Mezzanine level distribution is a likely solution to plant infrastructure issues. Under this scenario, video signals are compressed from 1.5 Gbps down to 200–300 Mbps using *higher level* MPEG-2 (higher level = less compression). SDTI is used as a carrier for the resulting compressed stream. This permits the continued use of the existing SDI plant infrastructure, which may consist of:

- Routers
- Digital-to-analog converters (DA)s
- Cabling
- DVTRs and servers (In this application, they must provide >200 Mbps raw data rate for playback and record).

In an ideal arrangement, new DTV devices would support mezzanine I/O. The data rate could vary from one device to the next, as it is assumed that encoders would support a variety of rates and be user-configurable. The higher the data rate, the better the quality of the compressed bit stream. The maximum data rate will be established based on the lowest of the following two parameters:

- The SDI router and available DAs (270 or 360 Mbps)
- The uncompressed *(bit bucket)* capability of the recorders being used

Mezzanine encoding/decoding is attractive because it is much less severe in terms of compression than 19.4 Mbps. Figure 5.10-8 illustrates one possible mezzanine implementation.

Audio Issues

The process of distributing digital audio through a video facility by multiplexing it into the digital video steam is not particularly new. Implementation roadblocks slowed acceptance at first, as considerations relating to the video signal were sorted out. SMPTE refers to this audio signal multiplexing process as *embedded audio.* AES-formatted digital audio signals are carried in the ancillary data space of serial digital video signals. This capability permits routing of video and accompanying multichannel audio through a single cable and through a single level of a routing switcher.

The benefits of this capability are obvious: it eliminates multiple cables, tandem audio/video routers and audio-to-video timing considerations when switching. Embedded audio holds particular advantages for a facility (or portions of a facility) that route or pass signals through without modifying the content. In other words, for strictly audio-follow-video applications such as uplink, STL and network feeds.

It is fair to note that timing considerations when using embedded audio do not disappear, they merely change form. The encoding/decoding process introduces delay into the audio element, just as it does

Table 5.10-2
Referenced Source Format Parameters.[2]

Reference Document	SMPTE 260M	ANSI/SMPTE 274M	ANSI/SMPTE 274M
Parallel word rate (each channel Y, C_R/C_B)	74.25 Mwords/s	74.25 Mwords/s	74.25/1.001 Mwords/s
Lines per frame	1125	1125	1125
Words per active line (each channel Y, C_R/C_B)	1920	1920	1920
Total active lines	1035	1080	1080
Words per total line (each channel Y, C_R/C_B)	2200	2200	2200
Frame rate	30 Hz	30 Hz	30/1.001 Hz
Fields per frame	2	2	2
Total data rate	1.485 Gbps	1.485 Gbps	1.485/1.001 Gbps
Field 1 EAV V = 1	Line 1121	Line 1124	Line 1124
Field 1 EAV V = 0	Line 41	Line 21	Line 21
Field 2 EAV V = 1	Line 558	Line 561	Line 561
Field 2 EAV V = 0	Line 603	Line 584	Line 584
EAV F = 0	Line 1	Line 1	Line 1
EAV F = 1	Line 564	Line 564	Line 564

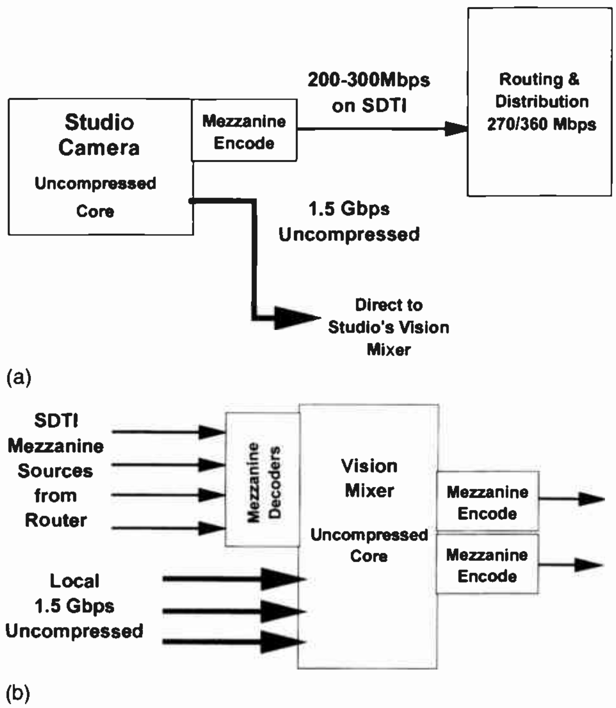


Figure 5.10-8. One possible studio implementation scenario for mezzanine level compression: (*a*) camera system, (*b*) vision mixer. *(Courtesy of Leitch.)*

with the video signal. For greatest flexibility, the serial digital system must be capable of delaying either the video or the audio to maintain synchronization between the two. These elements may necessitate additional expense and complexity, which must be balanced against the cost of additional audio levels in the routing matrix and a separate level of audio wiring within the facility.

Despite its intrinsic benefits, embedded audio rapidly loses its appeal in facilities where breakaway switching is required. When either the video or audio of a combined datastream must be changed, the two signals must again be separated. Time delay compensation must be provided for the delays introduced by this demultiplexer/multiplexer stage. The cost of the encoding/decoding hardware is, of course, also a consideration. Given these necessary tools, the desired changes can be made to the audio and/or video signal, after which the video and audio are recombined and sent on their way.

It is obvious that embedded audio is not usually practical in a post-production environment, except for intra-facility tie-line feeds and, perhaps, machine room switching systems.

Inside AES/EBU

Most professional digital audio devices employ an output protocol developed jointly by the Audio Engineering Society (AES) and the European Broadcasting Union (EBU). This serial digital transmission format, known as AES/EBU, has seen several incarnations. The basic format was agreed to in 1985, and soon thereafter, equipment began showing up built to the standard, known at the time as AES3-1985. Subsequent revision in 1992 led to the standard in use today, AES3-1992. Figure 5.10-9 shows the basic framing structure of AES3-1992.

The AES/EBS format transmits and receives left and right channel information using a data rate that corresponds exactly to the source sampling frequency. One *frame* consists of two *subframes*, labeled A (left channel) and B (right channel), each consisting of 32 bits of information. Each subframe, in turn, contains data for one audio channel. The first four bits are used for synchronization and for preamble identification. The next 4 bits are for auxiliary data. The next 20 bits carry audio data, with the Most Significant Bit (MSB) transmitted last; for 16 bit audio, 4 of the bits are set to zero. The last four bits form a control field consisting of:

- *Validity bit* (V), which indicates if the previous audio sample is error-free
- *User bit* (U), left uncommitted for a variety of user-related applications
- *Channel status bit* (C), which is used to form a data block. For each channel, one block is formed from the channel status bit contained in 192 successive frames
- *Parity bit* (P), which provides even parity for each subframe

Two interconnection methods are used for AES/EBU audio:

- Twisted pair cable with a characteristic impedance of 110 Ω and terminated with XLR connectors (balanced). This is the basic mode specified by AES3-1992
- Seventy-five ohm coaxial cable terminated with BNC connectors (unbalanced). This mode is the result of recent work between AES and SMPTE to accommodate the needs and available hardware of video users.

Interconnection between these two modes is possible with either passive or transformer-coupled matching networks.

Sampling rate

The AES/EBU specification is capable of supporting sampling rate frequencies ranging from approximately 25—54 kHz. Most equipment, however, will only operate at a given, specific sampling rate. Several standard rates have evolved over time; the most common being:

- 32 kHz, used sparingly in broadcasting, with the exception of certain satellite distribution networks
- 44.1 kHz, the standard for the compact disc audio system
- 48 kHz, the professional standard used by most broadcasters for original recording and transmission

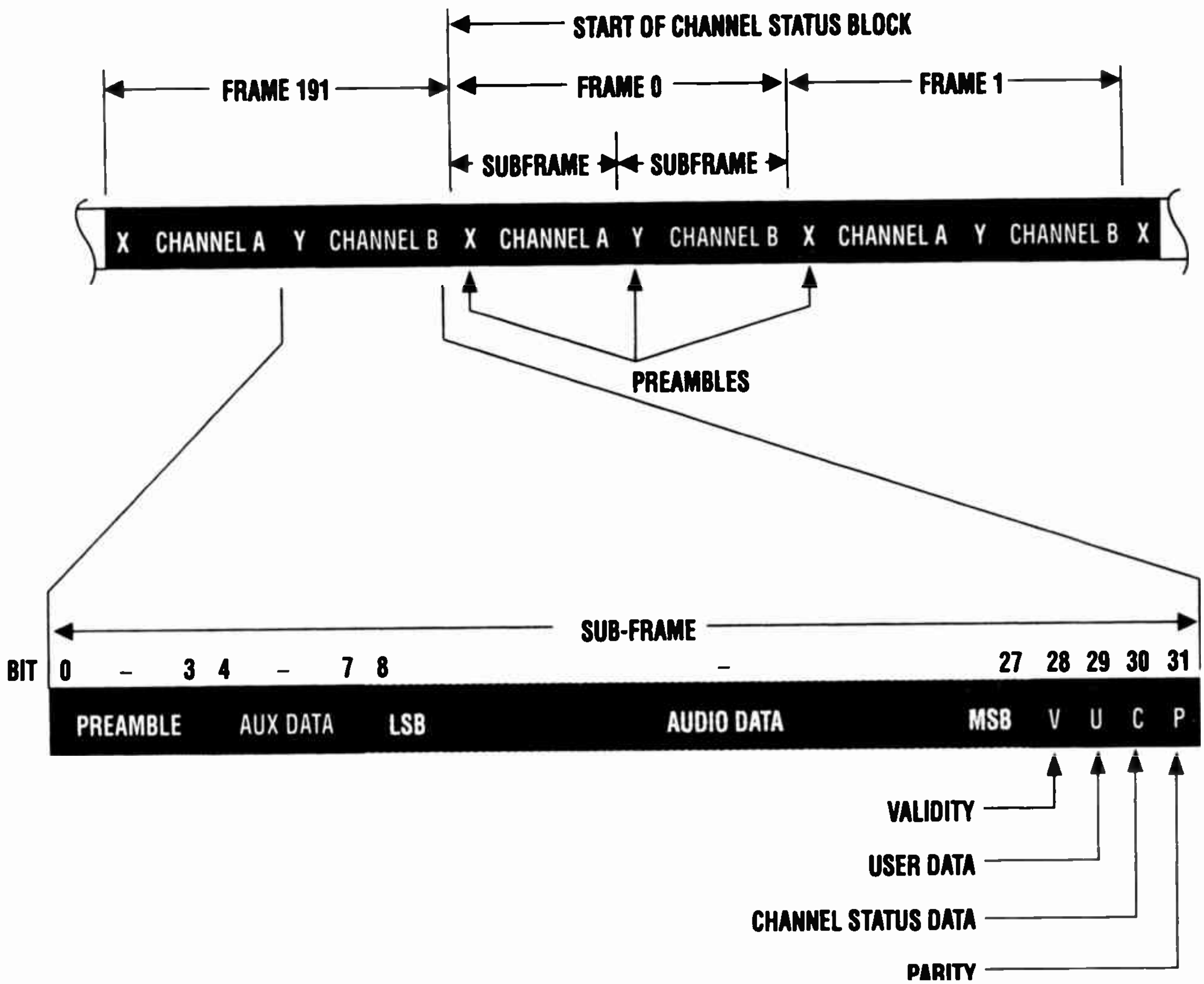


Figure 5.10-9. The structure of frames and subframes for the AES/EBU digital audio standard. (*Source:* AES3-1992.)

Although professional digital video equipment almost always uses a 48 kHz sampling rate for audio, TV stations and post-production facilities often use some quantity of CD-originated material in the production process, requiring them to deal with 44.1 kHz audio as well. Such situations require the use of a sampling rate converter to avoid unnecessary analog-to-digital (A/D) and digital-to-analog (D/A) conversions. It should be noted, however, that although the rate conversion process operates in the digital domain, it is not completely transparent and can have an audible impact on the signal.

Sampling rate synchronization is a related issue for application of the AES/EBU protocol. Audio and video digital devices have their sampling rates determined by an internal oscillator/clock. This leads, predictably, to sampling rate variations between machines operating at the same nominal sampling rate. For example, consider two digital audio tape (DAT) machines running with 48 kHz sampling rates and sitting side-by-side playing out material. It is possible—and perhaps likely—that the machines will drift with respect to a fixed 48.0 kHz reference; they will tend to periodically drift in and out of synchronization. Such a loss of synchronization between devices can cause audible artifacts, typically pops and clicks when switching or otherwise processing the signals. The worst case scenario is that the destination device will not recognize the incoming audio datastream. The result—dead air.

The obvious solution is to lock all of the audio equipment to a common reference, like the video gear. Most current technology digital video equipment derives its video and audio clocks and sample rates from the same oscillator or timebase. This being the case, the audio sampling rates should remain fixed and stable so long as the video equipment is referenced to a common source (either sync or composite black burst). There are, however, problems with extending this genlock philosophy to the rest of the facility, including:

- Most digital audio gear will not accept a video signal as a reference. The reference input (if provided) on digital audio equipment is typically an AES/EBU port or a *word clock* signal from another digital audio device. (A word clock is a square wave signal at the sampling frequency used to synchronize digital audio devices.)
- Video equipment will not typically accept any of the audio reference signals for its video reference.

As you might suspect, the relative phase of the sample clocks is also important for proper operation of an AES/EBU based digital audio system. The term isochronous is used to describe the locked but not phased condition; synchronous, on the other hand, describes two or more signals with coincident timing reference points. In order for an audio switch or mix to be clean and free of pops or other degradation, it must be made at the AES frame boundary, with the source and the destination devices having coincident frame boundaries and identical sampling frequencies. *Frame discontinuities* are the most common cause of pops and clicks in a digital audio system, and accurate timing between signals is the easiest way to eliminate them.

Asynchronous Transfer Mode

Asynchronous transfer mode (ATM) is a technology based on high speed packet switching. It is an ideal protocol for supporting professional video/audio and other complex multimedia applications. ATM is capable of data rates of up to 622 Mbps.

At the present time, ATM is primarily directed at telecommunications applications. Inter-facility video transfers over ATM are typically rather expensive and complex at this writing. Despite these current limitations, there is a great deal of interest in ATM technology for video and great promise for the future.

The Basics

ATM was developed in the early 1980s by Bell Labs as a backbone switching and transportation protocol. It is a high speed, integrated multiplexing and switching technology that transmits information using fixed length cells in a connection oriented manner. Physical interfaces for the *user network interface* (UNI) of 155.52 Mbps and 622.08 Mbps provide integrated support for high speed information transfers and various communications modes, such as circuit and packet modes, and constant, variable or burst bit rate communications.

These capabilities lead to four basic types of service classes of interest to video users:[3]

- *Constant bit rate* (CBR) which emulates a leased line service, with fixed network delay
- *Variable bit rate* (VBR) which allows for bursts of data up to a pre-defined *peak cell rate*
- *Available bit rate* (ABR) capacity is negotiated with the network to fill capacity gaps
- *Unspecified bit rate* (UBR) provides unnegotiated use of available network capacity.

These tiers of service are designed to maximize the traffic capabilities of the network. As illustrated in Figure 5.10-10, the CBR data streams are fixed and constant with time. The VBR and ABR systems vary as shown. The bandwidth of the UBR class of service is a function of whatever network capacity is left over after all other users have claimed their stake to the bandwidth. Not surprisingly, CBR is usually the most expensive class of service and UBR is the least expensive.

One of the reasons ATM is attractive for video applications is that the transport of video and audio fits nicely into the established ATM service classes. For example, consider the following applications:

- Real-time video, which demands real-time transmission for scene capture, storage, processing and relay, fits well into the CBR service class
- Non-real-time video, such as recording and editing from servers, distributing edited masters and other operations that can be considered essentially off-line, can use the ABR service
- Machine control and file transfer, such as sending still clips from one facility to another, find the VBR service attractive.

ATM is growing and maturing rapidly. It has already been implemented in many industries, deployed by customers who anticipate such advantages as:

- Enabling high bandwidth applications including desktop video, digital libraries and real-time image transfer
- Coexistence of different types of traffic on a single network platform to reduce both the transport and operations costs
- Long term network scalability and architectural stability.

In addition, ATM has been used in both local and wide area networks. It can support a variety of high layer protocols and will cope with future network speeds of gigabits per second.

The Cell Structure

It is worthwhile to explore the ATM channel format in some detail because its features are key to the usefulness of ATM for video. ATM channels are represented by a set of fixed-size cells and are identified

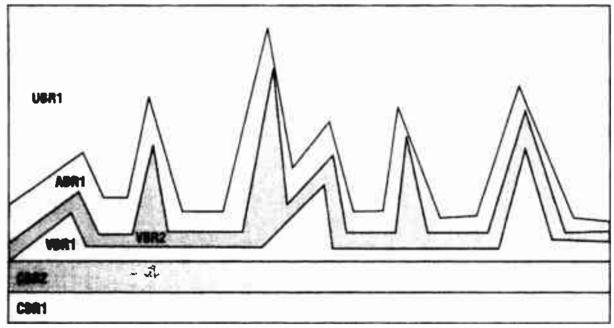


Figure 5.10-10. The typical packing of an inter-nodal ATM trunk.[3]

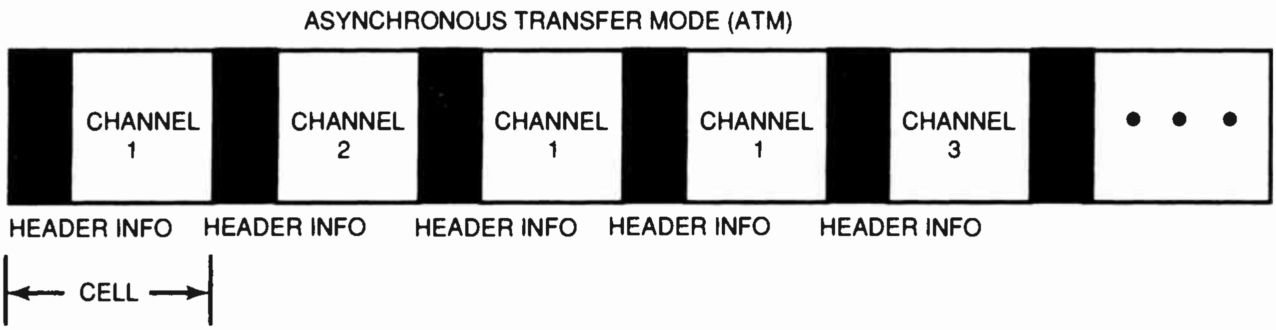


Figure 5.10-11. The ATM cell format.[4]

through the channel indicator in the *cell header*.[4] The ATM cell has two basic parts: the header (5 bytes) and the payload (48 bytes). This structure is shown in Figure 5.10-11. ATM switching is performed on a cell-by-cell basis, based on the routing information contained in the cell header.

Because the main function of the ATM layer is to provide fast multiplexing and routing for data transfer based on information included in the header, this element of the protocol includes information not only for routing but also fields to indicate the type of information contained in the cell payload. Other data is included in the header to perform the following support functions:

- Assist in controlling the flow of traffic at the UNI
- Establish priority for the cell
- Facilitate header error control and cell delineation functions.

One key feature of ATM is that the cells can be independently labeled and transmitted on demand. This allows facility bandwidth to be allocated as needed, without the fixed hierarchical channel rates required by other network protocols. The connections supported are either permanent or semipermanent, and do not require call control, real-time bandwidth management and processing capabilities; hence, the flexibility of ATM for video/multimedia applications.

Using ATM for broadcasting

Although ATM has been around for some years, it is still in its infancy with regard to video applications in general, and television broadcasting in particular. Generally speaking, ATM begins to look attractive today from a cost standpoint for point-to-point service over 50 miles in distance or for multiple site connectivity.

When additional network capacity and video interface systems are available and are more cost effective, it is quite reasonable to consider building a *virtual studio* where production elements can be physically separated and interconnected via ATM. This concept is illustrated in Figure 5.10-12. Note that feeds from the remote camera crew and studio are routed to a production suite, which also has access to an asset database. The finished product is then transported to the broadcast facility, where it is transmitted to viewers. Each of these five sites could be separated by many blocks or many hundreds of miles.

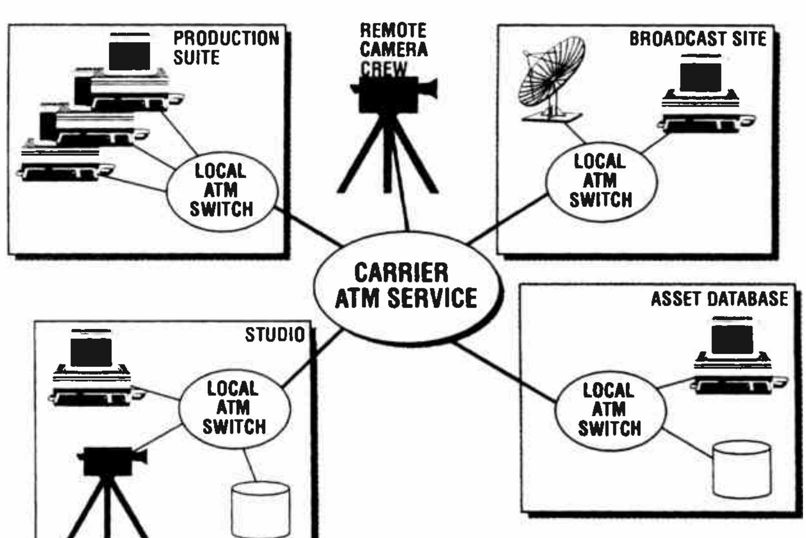


Figure 5.10-12. The concept of a virtual studio via ATM.[3]

REFERENCES

1. Fibush, David, *A Guide to Digital Television Systems and Measurement,* Tektronix, Beaverton, OR, 1994.
2. "SMPTE Standard for Television—Bit-Serial Digital Interface for High-Definition Television Systems," *SMPTE 292M-1996,* SMPTE, White Plains, N.Y., 1992.
3. Piercy, John, "ATM Networked Video: Moving from Leased-Lines to Packetized Transmission," *Proceedings of the Transition to Digital Conference,* Intertec Publishing, Overland Park, KS, 1996.
4. Wu, Tsong-Ho, "Network Switching Concepts," *The Electronics Handbook,* Jerry. C. Whitaker (ed.), CRC Press, Boca Raton, FL, pg. 1513, 1996.

BIBLIOGRAPHY

Pohlmann, Ken C., "Audio Recording Part 2: Digital Audio Technology," *NAB Engineering Handbook,* 8th edition, National Association of Broadcasters, Washington, DC, 1992.

Hunold, Kenneth, "Interfacing Digital Audio," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, pg. 80, April 1996.

5.11
INTERCOM AND IFB SYSTEMS

BRADLEY DICK
***BROADCAST ENGINEERING* MAGAZINE, OVERLAND PARK, KS**

INTRODUCTION

Intercoms. Everyone has used them, and probably without thinking about it. Put on the headset or press the button for instant communication. In today's complex and fast paced broadcast world, the need for rapid, reliable and flexible communications has never been greater. Fortunately, the available intercom equipment is capable of meeting almost any need that might arise. Unfortunately, the lowly intercom system is frequently left out of budgets or is grossly under funded. The key to the proper purchase, installation and use of intercom equipment begins with a thorough understanding of the equipment basics.

The historical background of intercoms, as far as broadcast is concerned, lies primarily in television. The early radio stations had less need for intercoms because of the ability to use hand signals. In television, it was seldom possible for the camera operator to see the director, let alone take visual hand cues. In fact, it was the need for TV directors to be able to communicate with camera and floor personnel that brought about the first broadcast intercom systems as we know them today.

From a usage standpoint, there are two basic kinds of intercom systems: point-to-point (matrix) and conference line. These names stem from *how* the communications take place; either in a point-to-point, (private line) manner or on a conference (also called party line, or PL) where everyone hears everything. To further confuse matters, digital technology has resulted in a new classification for intercoms, although they do the same thing. From an overall design standpoint, these new digital systems are sometimes are referred to as distributed crosspoint and centralized crosspoint systems. And if that is not enough confusion, the term digital system may only refer to digital control or the system may be fully digital control and audio. Related communications systems called *IFB* and *ISO* are often included in any intercom discussion.

SYSTEM ARCHITECTURE

Point-to-point or *matrix systems* allows communication with a selected person or area. The communication travels from one point to another point. No other people or locations receive that communication. These types of intercoms provide complete privacy between stations and may also provide custom programming capability from a central computer or terminal. With many systems, groups or *virtual party lines* can be created to provide great flexibility.

Matrix systems require 4-wire interconnection (although a 2-wire matrix is possible). Such systems are well suited to post production operations where the system configuration remains relatively constant. In a 4-wire topology, talk and listen signals to and from each station are carried on separate paths. These paths may be wires, RF, digital links or fiber optic paths. Matrix systems usually allow stations to selectively call other stations and there is generally a tally indication at the receiving point showing who is calling.

Historically, point-to-point or matrix system consisted of a centralized card frame and signal routing equipment controlled from remote stations. Today, distribution technology affords more versatile equipment arrangements. The audio signal paths are typically analog and simplex.

Each station in a matrix system typically interconnects via a 4-pair cable. Two pair of conductors are used for audio receive and listen for both the main and remote station and one or two pair of conductors are used for data (typically RS-422). Some of the newest digital matrix intercom systems operate on a single coax cable.

The system allows a station to route its voice to one or more other stations. Normally, point-to-point systems require direct (home run) cabling from each station to the central rack. In this system, the speaker decides who hears the communication. The listener normally has no control over who is received at the individual stations.

Some advantages of 4-wire matrix systems include: simple interfacing to other devices, no hybrid null problems, selective calling to all stations, call tally and privacy. Disadvantages of the 4-wire topology used to include all the multipair cabling required. Modern matrix systems, however, eliminate most of these problems.

The second type of intercom system is the conference line system. Also known as party line (PL) systems. These are the simplest form of intercom system, typically used for teamwork activities like remote productions. Anyone connected to the intercom line hears all of the communication that takes place. When someone speaks, everyone hears it. Live studios and outside broadcast activities are best served by PL systems.

PL intercoms are generally 2-wire systems. Talk and listen audio is carried on the same pairs. Power may be carried on the audio pair or separate conductors. A party line intercom can also be implemented with 4-wire techniques, but it is much simpler and less expensive to do so with the 2-wire technique.

In a PL system, all operators share a common channel or number of channels. The only privacy is what may be provided by multiple channels or a camera isolation subsystem. Some signaling can be provided, but usually not to the extent that every operator can selectively signal another operator.

The conference line (PL) system is sometimes called a *distributed amplifier or distributed crosspoint system.* This means that each station is equipped with all of the required electronics for both receiving and transmitting audio and call signals. Conference line systems require minimal centralized rack equipment, which typically consists of the system''s power supplies and passive assignment switching in multichannel systems.

Conference line systems allow groups of stations to communicate in real time, full duplex fashion. Multichannel conference systems allow users access to several different channels, determining who they talk and listen to. Keep in mind that with the conference line system all stations on a particular loop hear everything on that loop. Normally, there are no private communications such as point-to-point systems provide, although multiple *channels* may be provided. (Note, channels and stations are not the same thing, see Key Terms at the end of this chapter.)

Advantages of 2-wire PL systems are: simple wiring, easy expansion for additional stations, less complex central equipment, low cost per station and simple operation, which is important for team activities like remotes. Of prime importance, 2-wire PL systems can be quickly and easily reconfigured, which is especially important for remote trucks and mobile venues.

There may be a size limitation for 2-wire PL systems, especially in terms of ease of installation. Once you go above a five or six station system, the matrix solution may become more viable—and easier to use.

Some disadvantages of 2-wire PL systems are their reliance on hybrids and the difficulty in interfacing them to other (often incompatible) 2-wire systems, lack of selective calling and limited privacy. Figure 5.11-1 shows a typical single channel conference line system.

It may seem that point-to-point and conference line systems are mutually exclusive. In fact, it is not uncommon to find a combination of both systems. The more sophisticated point-to-point and conference line systems currently available usually have the ability to integrate both types of communications functions within one system. This usually means that multiple channels are available. Stations can be assigned to different channels for communication, but communication between channels is minimal.

(Table 5.11-1 summarizes the relative merits of the various architectures.)

Number of Wires

As discussed earlier, intercom systems are also sometimes classified by the number of wires required for the audio path, not including any conductors for dc power. If it takes two wires to carry the audio, it is called a 2-wire system. If it takes 4 wires to carry the audio, it is then called a 4-wire system. There are also references to 3-wire systems.

With a 2-wire system, full duplex communications can generally take place on one pair of wires. This means that conversation travels in both directions on the same pair of wires. Some 2-wire systems run balanced intercom audio, others run unbalanced audio. Also, some intercoms put the dc operating voltage on the same two wires. Others require one or more additional wires for dc power or use a local power supply. An example of a 2-wire system combining both audio and dc power on a single pair is the telephone.

In a 4-wire system, one pair of wires is used for each simplex (one direction) path of communication. Therefore, two pairs a transmit pair and a receive pair (four wires) are required for the complete two-way communication link. Generally, 4-wire systems require a local dc power source or additional conductors for dc power, plus the interconnect.

Defining 3-wire systems is confusing, as there are several different ways 3-wire can be interpreted. It can refer to equipment and systems that interconnect using standard 2-conductor shielded microphone cable terminated in 3-pin XLR type connectors, regardless of how many conductors are required for the audio path. (Two conductors plus the shield equates to three wires.) The equipment of most manufacturers of conference line intercom systems interconnects via this standard microphone cable. But different systems use the microphone cable in different ways.

One technique places one channel of unbalanced intercom audio between one conductor and shield and dc operating power between the other conductor and shield. (Because this type of system normally requires all three conductors to operate, it is sometimes referred to as a 3-wire system.) This is the easiest conference line system to switch and route.

A second design provides two channels of unbalanced intercom audio, one channel between each conductor and shield, and combines the dc operating power with one of the intercom audio channels (see Figure 5.11-2). Since this type of system actually only requires two conductors to operate, it is sometimes referred to as a 2-wire system, despite that it interconnects with 3-conductor cable.

A third design provides one channel of balanced intercom audio between the two inner conductors with the dc operating power phantomed between the shield and the inner conductors. If two conductors are used, this would only be a one channel system. (Again, because this type of system actually requires all three conductors to operate, it is sometimes referred to as a 3-wire system. Only one manufacturer uses this method.)

Three-wire can also refer to the tip, ring, sleeve (TRS) plug or jack typically used on the old carbon

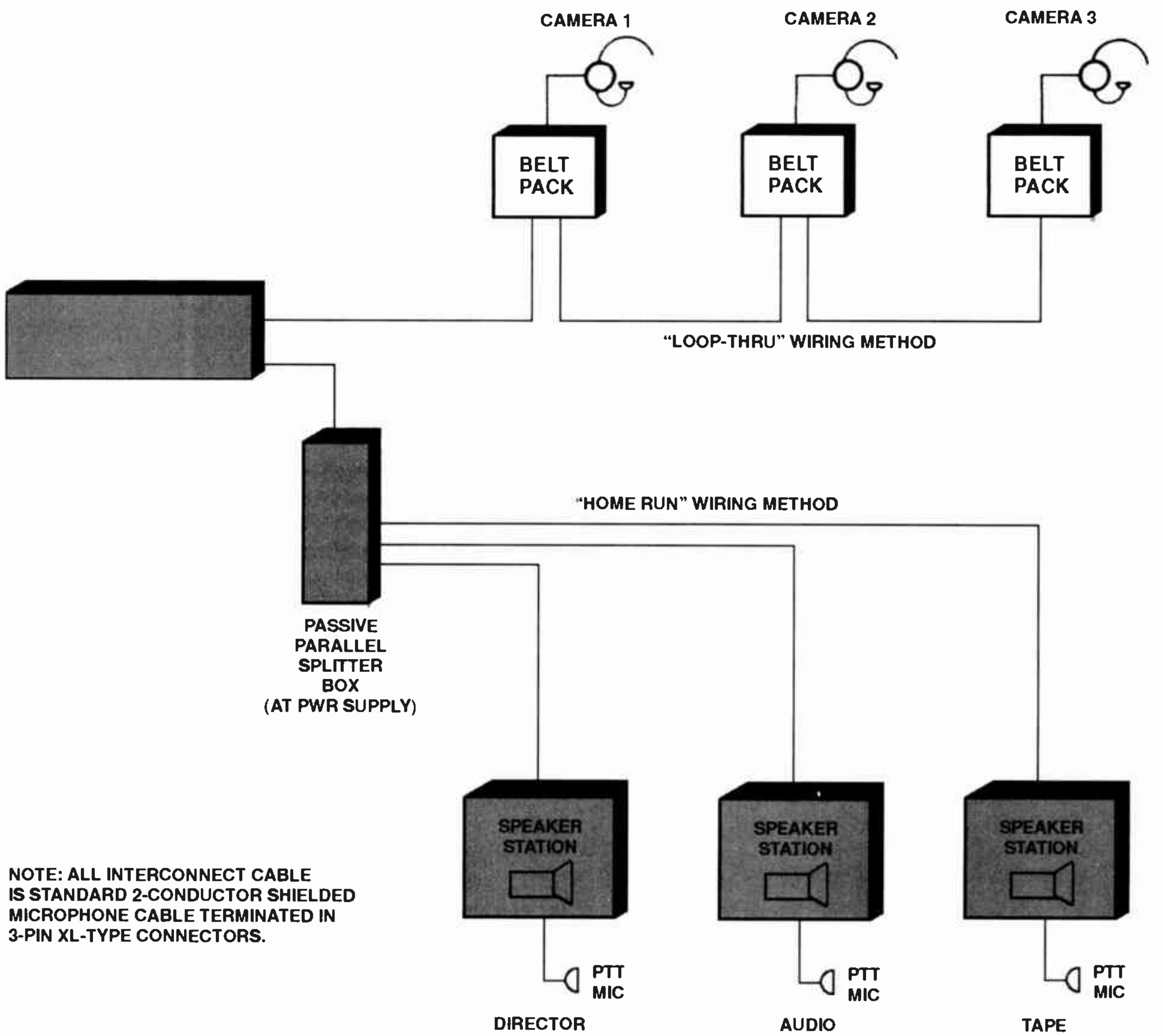


Figure 5.11-1. A simple 6-station single-channel conference line intercom system. (Courtesy *Broadcast Engineering*). Note: All interconnection cable is standard 2-conductor shielded microphone cable terminated in 3-pin XLR-type connectors

mic type headsets. This is essentially a 4-wire circuit with one side of the headphone circuit and one side of the microphone circuit tied together in common, making it a 3-wire system. (Note: Any headset is inherently a 4-wire device, two wires in the receive circuit going to the earphone and two wires in the send or transmit circuit coming from the microphone.)

The Need for Interfaces and Power

Regardless of what's happening at the central switching matrix, certain things must still happen at the business end of the intercom. Intercom stations are powered in two ways. Rack or wall mounted (fixed) stations can be powered from available ac by way of external dc supplies. These stations can then communicate via dry lines carrying audio only. Beltpack stations, on the other hand, are powered via wet communications lines, carrying 24VDC to 32VDC and audio. Some fixed stations can also take power from a wet line.

However supplied, the power is used to drive several necessary functions. With the low impedance headsets in use today, intercom stations must provide a headphone amplifier with volume control, switching circuits for microphone and channel selection, call light power and switching, and a preamp for the headset microphone and sidetone amplifier usually with screwdriver-adjustable gain control.

It is important to remember that not all intercom makers get power to their beltpack in the same way, even though they all use the same standard low impedance microphone cable and XLR connectors. A case in point is the difference between the 2-wire schemes

Table 5.11-1
Comparison of 2 and 4-Wire Systems

2-wire (party line, conference) system advantages:	2-wire (party line, conference) system disadvantages:	4-wire (matrix, point-to-point) system advantages:	4-wire (matrix, point-to-point) system disadvantages:
Simple wiring.	Reliance on hybrids.	Simple interfacing to other equipment systems.	Requires lots of central equipment (rack space).
Easy to install and expand with additional stations.	Difficult to interface to other manufacturers' 2-wire systems.	No hybrid null problems.	May require more time to install and requires more wiring.
Little central equipment required.	Difficult to implement selective calling.	Easy to interface to audio systems.	May be more expensive than other solutions.
Low cost per station.	Limited privacy capability.	Customer programming capability.	May be difficult to expand, expansion often require the addition of blocks of stations, rather than individual stations.
Generally less expensive than matrix (4-wire) systems.		Selective calling capability.	Analog systems require significantly more wiring than conference line systems, but digital technology can reduce this.
Offers a wide variety of stations and user features.		Tally usually provided.	Control panels may be larger and cumbersome to operate, especially if the system has a large number of stations.
Best choice for remote production units.			Offers limited flexibility, especially for mobile applications.
Simple operation for team type activities.			Generally used for large, in-plant applications.
			Provides instant, unblockable communications to selected points.
			Inappropriate for field use.

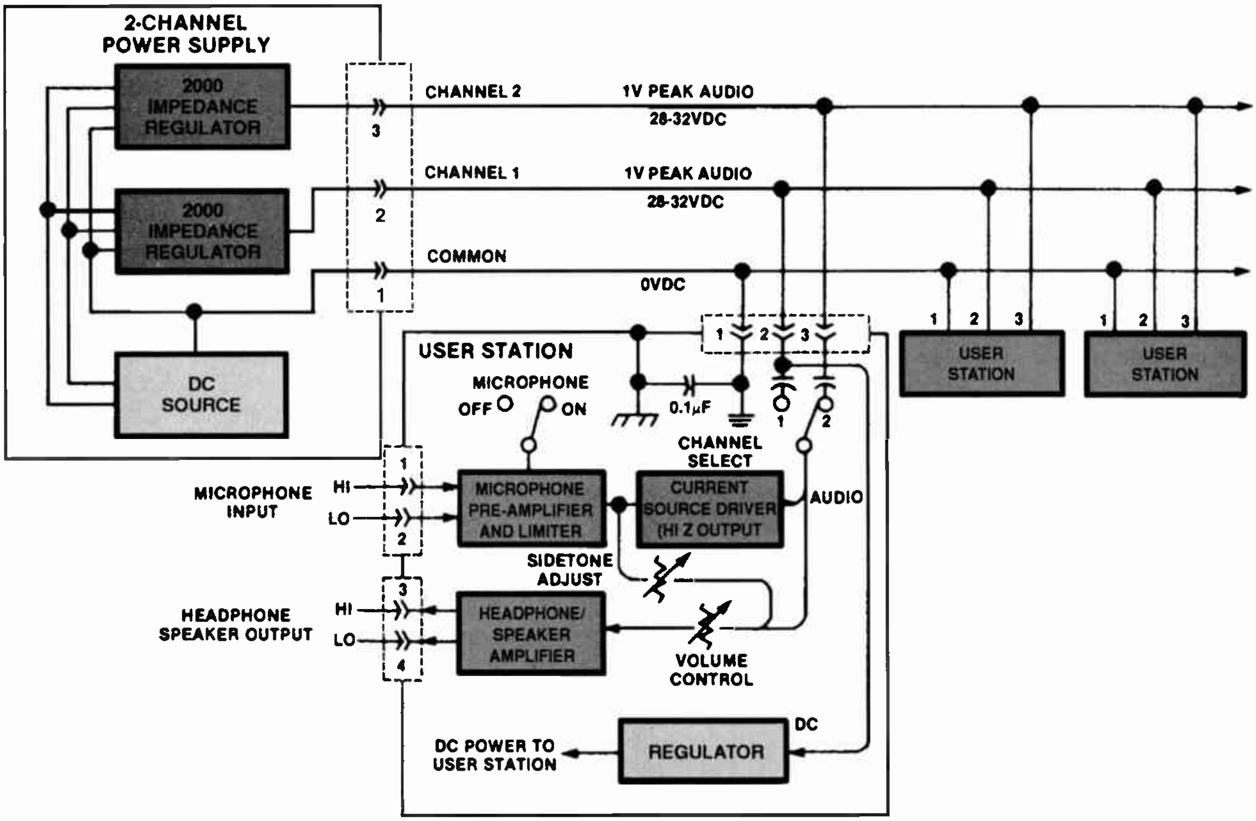


Figure 5.11-2. A 2-wire intercom system can provide two channels of communication on a single standard microphone cable. In this example, the dc power for the system carried along with the channel one intercom audio. (Courtesy *Broadcast Engineering*)

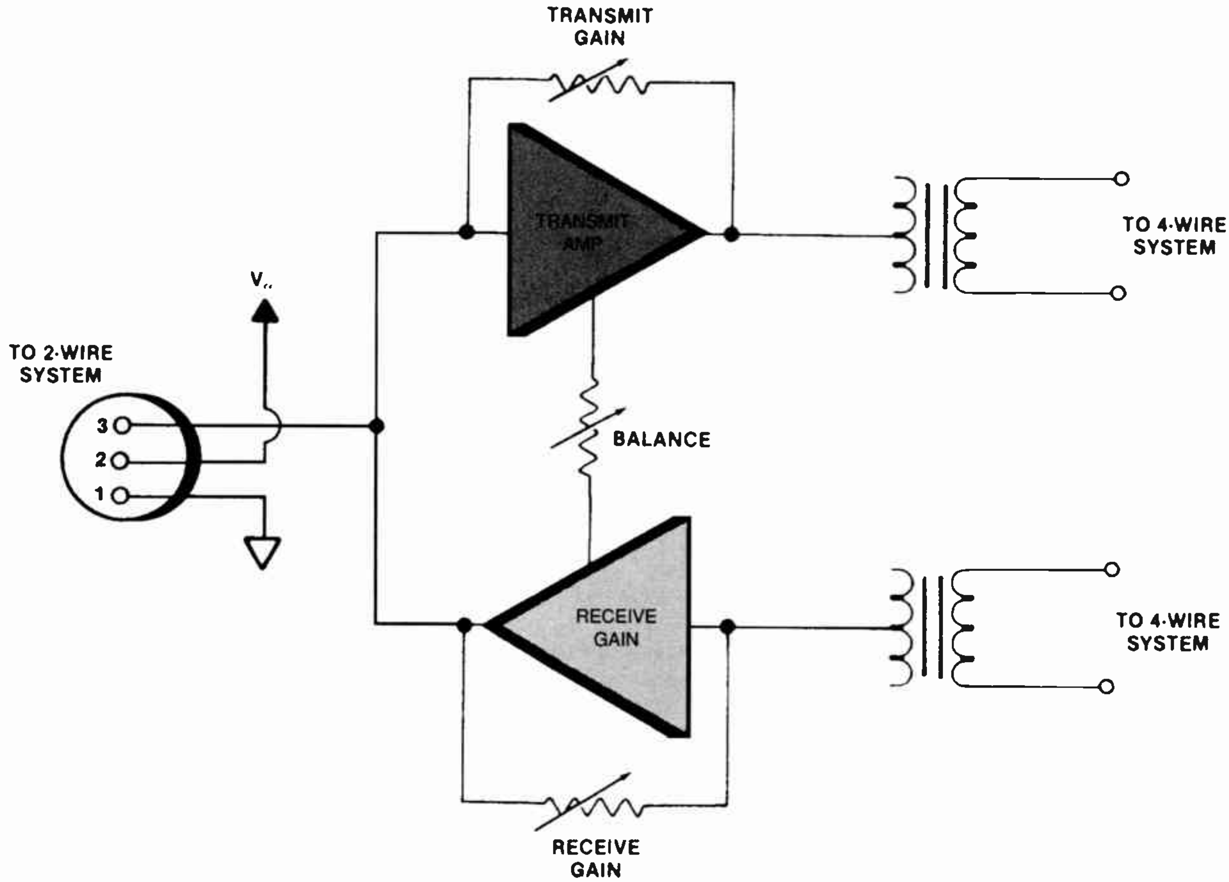


Figure 5.11-3. Connecting a 4-wire intercom signal to a 2-wire intercom system is easy if a standard 4-wire to 2-wire interface box is used. The box provides all of the necessary impedance and level correction. (Courtesy *Broadcast Engineering*)

used by Clear-Com and Telex. Clear-Com uses pin one for ground, with operating voltage on pin two and audio on pin three. Telex uses a method similar to microphone phantom powering, using pin one for ground and pins two and three for both balanced audio and simplex power. Figure 5.11-3 illustrates how to interconnect a 2-wire intercom to a 4-wire intercom.

Call lights on these systems also operate differently. RTS/Telex systems use a 20 kHz tone while Clear-Com uses a dc voltage for signal activation. These methods are not compatible, and if there is ever a situation where a remote truck with one system pulls up to a venue equipped with the other, an interface will be needed. Fortunately, such adapters and interfaces abound. Almost every manufacturer makes an adapter for its gear to plug and play with most other makers' equipment. Any well-equipped remote truck should have such a universal interface, which is typically packaged as a rack mount unit.

These interfaces generally provide bidirectional, 2 channel communication between the hosts system's and the foreign system's 2 channel lines. The interface usually can be powered by either wet lines or via local dc. The lines on each side of the interface are assigned just as if they were beltpacks and each side of the interface can be powered by the wet line powering scheme to which it its native. These units are versatile and provide the necessary impedance, level matching, pin conversion and call signal translation. This is certainly an improvement over the lighting director having to wear two headsets(one to the light crew and one to the TV truck.

Finally, some 2-wire base stations use different connections for single channel beltpacks versus 2 channel beltpacks. If single channel beltpacks are mixed up with 2 channel connectors, the whole system can be loaded down. This is usually revealed by a sudden drop in reception level at every station. Beltpacks on these stations are usually color coded to differentiate between single channel and 2 channel units. It pays to also color code the corresponding intercom connections on the studio truck breakout panel because the single and dual channel units once again use the same type of cable and 3-pin XLR connectors.

Camera Systems

Camera intercom systems create the biggest problems with interfacing and providing quality audio signals. Today's production environments often require lots of headsets be coupled onto the camera line. However, simply tacking additional headsets onto the camera intercom line will not work. Typical camera intercoms provide very little audio power to the headset. This level is frequently insufficient to drive some headsets, especially if the ambient noise at the camera's location high, such as a live sporting event. The solution may be to use an external beltpack intercom station, which typically can provide up to 2 W of audio power, more than enough to overpower any ambient noise.

Unfortunately, multiple standards seem to be common with camera intercom systems. Furthermore, manufacturers are not consistent with intercom design within their own products. It is not uncommon to find that a single line of cameras from one manufacturer has completely different intercom systems. Therefore, providing good communications sometimes means modifying the camera intercom. Four basic approaches can be taken to improve the camera's intercom.

The first is to completely ignore the intercom. A separate cable can be tagged along the camera cable and coupled to beltpacks and headsets. This approach has the advantage of providing communications that are equally as good as the rest of the plant"s intercom system. This also has the advantage of guaranteeing compatibility between intercom stations. The disadvantage is that another cable along has to be run along with the camera cable and separate beltpacks and headsets must be provided.

A second method of dealing with the camera intercom also requires that separate wiring be used for the intercom. If the camera operates on multicore cable, the conductors normally used for intercom and program audio in this cable can be used for the external intercom. This modification involves bypassing the internal camera intercom totally, and bringing the circuit out of the camera head for connection to an external beltpack. While it does eliminate adding a separate cable to the camera, it means modifying both the camera and the Communications Control Unit (CCU).

Unfortunately, this option will not work with triax cable. Also, crosstalk and noise between the video and intercom circuits can sometimes become a problem if the conductors used for the intercom have insufficient shielding. For instance, using the camera's ground can result in horizontal sync buzz in the intercom. Try for a separate ground path here whenever possible.

The third method would be to completely remove the camera's internal intercom. In this case, the camera's intercom circuits are replaced with a standard intercom card from the same manufacturer as the external system. This method has the advantage of providing good audio quality and compatibility with the remainder of the facility's intercom. Also, this approach works with both multiconductor and triax cabled cameras. One disadvantage is that not all cameras have sufficient space for the additional circuit card. It also requires that all cameras be modified. An unmodified camera may not work with the modified intercom system.

The fourth method of modifying a camera intercom system is the easiest, but perhaps the least desirable. In this case, an interface device is placed between the existing camera intercom and the facility intercom. For optimum results, each camera should be interfaced individually rather than collectively. The interface box, essentially an active 2-wire to 2-, 3- or 4-wire hybrid, provides the necessary conversion of line and levels. Because there may be a significant difference in levels between the camera and control point, optimum results are not always possible. There is a limit to the amount of level correction and isolation that can be provided with the interfaces. This approach has the advantage that no modifications are required to the camera or its CCU. The disadvantage is that the audio levels and quality are limited by the original camera circuits. The interface box cannot always make up for poor levels from the camera.

Wireless Intercoms

Generally, there are three types of wireless intercom systems. The first provides a one-way, listen-only feature for the remote stations, typically an interrupted feedback or foldback (IFB). In this system, the master transmitter is tied to the intercom line. All communications on the intercom line are relayed to all of the wireless receivers. This type of one-way intercom is typically used for people who need to know what is going on, but do not need to talk back.

The second is a two-way system. In this configuration, the base unit and the field units can both talk and listen to each other. The two-way wireless intercom system can be further broken down into three sub types. The first is a simplex, 1 frequency system, as shown in Figure 5.11-4a. All stations talk and listen on the same frequency (F1). Everyone uses a push-to-talk (PTT) switch so users cannot listen and talk at the same time. If the system is tied to a conference line, it must be turned off before any field units can talk back to the base or other units. In this configuration, even a 6 station wireless system requires only one frequency.

It is also possible with several intercom systems to incorporate a two-way transmitter (Land Mobile-type) into the communications link. An interface is used to connect the two-way transceiver to the intercom system. Signalling from the intercom link can then be used to activate the transmitter. This technique is especially useful in situations where broadcast sites are located so far apart from each other that wired interconnection is not convenient or cost prohibitive. If repeaters are a part of the two-way system, even wider intercom range operation can be obtained.

The second type of two-way intercom system is a partial full duplex, 2-frequency system (see (Figure 5.11-4b). The units still use a PTT switch. The base station is tied to the conference line and is always transmitting to the remote units (F1). The remote units can then hear all of the communications on the conference line. All field units talk back to the base on the same frequency (F2) so only one unit can talk at once. This approach requires only two frequencies, no matter how many stations are involved.

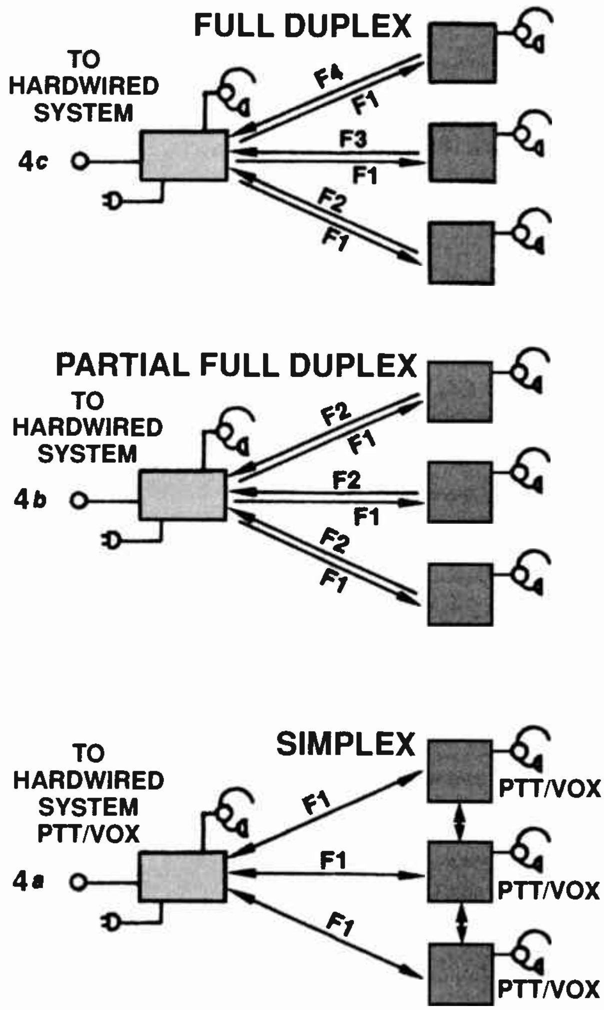


Figure 5.11-4. Wireless intercoms provide complete portability for the user. A full duplex system, shown in 4c, requires one frequency for each remote station, plus one frequency for the base station. (Courtesy *Broadcast Engineering*)

The third, type of two-way wireless intercom is more complex, but also much more flexible. It is also more costly. In this full duplex design (Figure 5.11-4c), everyone receives on the same frequency (F1), as described previously. However, each of the field units transmits back to the base on different frequencies (F2, F3, etc.). In this configuration, any number of remote units can listen and talk at the same time. A system with six remote units requires seven frequencies. This design is really a wireless version of a wired conference line intercom and no PTT switches are required. In a high quality system of this type, a wireless remote headset is virtually indistinguishable from a wired headset.

Virtually all wireless units will interface with wired systems either by themselves or through interface boxes. However, take care to not connect wireless intercom transmitters directly to an IFB system using wet lines. The powering on these lines can damage the transmitter. Instead, buffer the feed to the wireless transmitter using hard-wired IFB talent beltpack that outputs only audio on a dry line.

RF base units are usually generic in that they can operate with a hard wired system from the same manufacturer or they can be internally dip switch programmed to work with other intercoms. All generally offer a real panel loop-through just like hard wired beltpacks. Many models have two antennas on the base and two on each beltpack. This does not imply any type of spatial diversity as in wireless microphone systems rather, one antenna is used for transmission and the other for reception. Diversity reception in not commonly found on wireless intercoms.

Because intercom beltpacks not subject to the same size constraints as wireless microphones, they typically operate on a RF output power of 50 mw, or about five times the RF power output of many wireless mics. These systems have traditionally used the VHF band, but as with wireless microphones, many users are moving to the UHF band because of overcrowding and interface at VHF frequencies.

Wireless base station front panels generally include all the necessary connections: headset indicator light for continuous base transmit and a light to denote reception of each beltpacks' RF carrier.

The advantage of the wireless system is the mobility it provides users. The potential disadvantages include limited range, the possibility of interference and dropouts. Proper frequency coordination at the site in which the system will be used is extremely important. Failure to do so could result in a complete communications failure if a stronger RF signal appears in the area during the broadcast or taping. With the advent of DTV, frequency coordination is even more critical.

New digital solutions to the wireless intercom have recently appeared. One provides the option of up to two channels of frequency agile operation in either the VHF or UHF band. Another model provides the ability to switch between two channels of transmit and two channels of receive on each unit.

For private applications, like major league sports, an encrypted version is available. Communication takes place like on other wireless units, but the audio is digitally encrypted. Providing more than 64,000 different settings, eavesdropping is effectively eliminated.

An example football system might include: two channels of offense duplex communication (offense 1 and offense 2), one channel for defense and one channel for special teams. Connecting the system into four channels of a traditional wired intercom channels would provide extremely versatile operation. The head coach, using the two channels of his intercom, could switch between channel A (offense 1) and channel B (the defensive coaches). The offensive coaches have two channels, split among their three coaches. The defensive coaches area all connected to one channel and do not communicate with any other channels, but can still talk to the head coach. The special teams coach uses one channel and does not communicate with other coaches.

An encrypted wireless UHF intercom system can cost about 10 times what a wired system will cost. An non-encrypted VHF system will cost about five times that of a wired system.

SIGNALING

As mentioned previously, signaling is an important aspect of intercom capability. In a conference line intercom, signaling is normally referred to as a call or signal. Signaling provides a means (usually visual) to attract the attention of a person who is not wearing a headset or has turned off the loudspeaker. In a conference line system all stations in a given channel of communication will be signaled at once. It is generally not possible to signal individual users in the conference line system. Although signaling is usually a visual indication, it can be made audible for special situations.

Signaling is accomplished by applying either a high frequency tone or a dc signal to the audio pair. When using audio instead of dc for the trigger, the tone passes through any interfaces that may be in the line. However, dc signaling is less complicated and usually costs less. Recall from before that signaling can be used to activate a two-way transceiver for longer distance communications.

Signaling simply to get someone's attention has limited application in teleproduction. It is more often used in theater work or industrial applications. There are, however, many unique applications for signaling in teleproductions that extend beyond the mere calling of someone to the station.

Signaling functions of the intercom can also be used for remote control purposes. For example, in a small ENG or SNG truck, the call light signaling circuit can be wired to turn on the two-way radio through a relay, operating in a PTT mode. When the signaling is activated, the talent and engineers know that the studio is hearing the conference line audio because all of the signal lights are illuminated.

Another use for signaling is for push-to-talk-and-listen features. One system allows the signal circuit to be tied to the microphone circuit of a remote station. As an example, a director presses the push-to-talk button to address the tape room. In this case, the videotape operator does not have to press a switch to respond to the director. The signaling circuit at the remote station automatically turns on the microphone in the tape room, allowing hands-free communications from the tape operator. The signaling circuits can also be used to override a volume control on a paging speaker or in conjunction with IFB audio.

Conference line systems that provide signaling as a standard feature can be thought of as providing a free remote control capability. The applications for signaling/remote control functions are limited only by the engineer's imagination.

Program interrupt and ISO

Although not strictly a two-way intercom, IFB is often an integral part of a facility's intercom system. Also called program interrupt, or PI, director interrupt, director cue, IFB is used to cue both on and off air talent.

The talent end of the IFB system usually takes one of two forms. A subminiature in- or over-the-ear earset is often used in news and field report situations. Sportscasters often use dual earmuffs, one of which has the IFB channel.

The program audio is fed to the talent, usually through an in-the-ear earphone. Then, the program audio can be interrupted or attenuated by an operator (usually by the director or producer) to give verbal cues to the talent. This allows the talent to continue to talk, while at the same time listening for directions on future actions. IFB systems can be quite extensive involving multiple program inputs and links to off premise locations. Their heaviest use is in sports and news operations.

There is a less well known application for IFB in non-broadcast uses. In this application, the talent wears a small receiver tuned to the base station frequency. The audio output of the receiver is connected to an inductive loop antenna worn around the talent's neck. Finally, an ultraminiature receiver/earphone is worn in the talent's ear. The receiver/earphone is so small that it cannot be seen by the camera or audience and picks up its signal from the loop antenna. With this equipment, the talent can receive instructions from a director or lip-sync to an audio track. Stunt people often use this type of arrangement to coordinate their work on complex scenes.

Most IFB systems contain more than one channel (typically four or more) and allow selective communication to individual talent without disturbing others. Systems typically provide tally signals when there is more than one control point. In some systems, the program audio is cut off for the duration of the announcement or cue.

Other designs merely duck the program audio under the announcement. This latter approach adds a confidence factor for the talent in that they can still hear the program signal during the cues, but at a reduced level. IFB systems are available either as stand-alone systems or as an integral part of an intercom system.

An *ISO*, which means *station isolate* is like IFB, but the communication between the caller and destination can be two-way and private. A typical application would be for studio or remote camera intercoms where all operators are normally talking on a common party line. However, each camera intercom station can be interrupted, usually by the TD or director, and private communication take place. The communication removes any IFB audio and any other stations for a clean, private feed between the two points. This feature is typically available in the new matrix type intercoms.

REASONS TO UPGRADE A SYSTEM

There are a lot of reasons to upgrade to a newer intercom technology. Sometimes it is simply because the number of users has grown beyond the system's capabilities. Other times, additional functionality is needed. In general, there are some often heard reasons for installing a better intercom system.

- Client driven. Sometimes a complicated show requires more operators with additional user stations. Or, as the hours of operation within your facility expand, scheduling/equipment use eliminates the time sharing that was once possible with the existing system.
- Having to interface with other new equipment. New cameras will not interface with the current intercom system. As the industry races ahead, fueled by new technology, some otherwise acceptable systems will have to change to accommodate the newer hardware.
- The current system is too small and can not be expanded due to manufacturing limitations. Even if the inevitable end of the useful life of a system has come, do not trash it. Some of the older items can probably be reused and interfaced or updated to work with the new technology.
- The time and effort spent reconfiguring the current system to meet the needs of regularly schedule events is too great.
- Finally, as engineering becomes more of a freelance task, especially in maintenance engineering, a facility may need to streamline its functionality. Reliability is the name of the game with intercoms. A system that has all the necessary functions all the time is more reliable and consistent than one that is built as the needs arise.

Planning Your New System

Planning for a new intercom system should center on a careful step-by-step analysis of the station's (or truck's) communication needs. Purchasing an intercom is similar to buying a computer. First decide what the operational requirements are going to be to consider the hardware.

It is easy to underestimate the time and effort it takes to design an intercom system because the subtleties and requirements of the various features and systems are difficult to identify. An intercom system is a foundational element in a broadcast facility and if the proper choice is not made early on, other decisions get more confusing. Start early in the overall project with a decision so that other equipment choices will be less complicated.

Probably the most important knowledge the system specifier must have is an accurate understanding of the operational routine of the facility. This is hard to come by in a new facility or in a facility that is moving to new quarters. It is slightly less daunting to upgrade a current system because the work habits and operational routines are easily identified.

A facility-wide intercom system, be it an SNG truck, news control room, production facility or a network master control, is typically a combination of multiple types of intercom systems, some tied together, some stand-alone. While in theory an intercom system is either a PL or matrix design, the final configuration will usually encompass elements of both identities.

For instance, the ability to talk to on-air talent may be a completely separate intercom system from the one used by engineering to setup a remote truck incoming feed. Likewise, the master control operator might have a microphone on the console that simply pages to a speaker in the tape machine room. Again, the key is to thoroughly understand the needs of the facility.

Intercoms can be as simple as two people in different rooms communicating, such as a video editor and a tape operator, or as complicated as a NASA mission control room with users located all over the world and beyond. As an intercom system grows in size, so does the need for focused attention from the engineering department.

Digital Solutions

No look at intercoms would be complete without a discussion about digital intercoms. The term *digital system* and *digitally controlled system* are used interchangeably today. The key point is that digitally controlled or digitally based does not necessarily mean that the signals between stations are digital. It may mean that the central switching matrix and audio processing is digital. Although several current intercoms feature a digital switching matrix or internal digital audio processing, there are only a few on the market that actually send digital audio signals to the user stations. Various digital implementations are available from Clear-Com, Drake and RTS/Telex. These digital systems also support both new digital rack mount stations as well as the traditional analog stations (see Figure 5.11-5). All beltpack and earpiece stations on these systems still operate in the analog domain.

These new digital systems allow remote key panels to be located almost anywhere. Using dark fiber, coax or even an ISDN line, it is possible to install an intercom control panel in one city and have complete communication capability to another location thousands of miles away.

Note that both digitally controlled and fully digital systems have interfaces that provide the conventional 2-wire analog lines to beltpacks, but not all beltpack wiring schemes are compatible—even though they all use standard microphone cable and XLR connectors. This is an area where operator knowledge is critical. Beltpacks are not interchangeable and using the wrong type can load down the system or even cause damage.

One feature that totally digital systems have yet to implement is the ability to daisy-chain stations like current analog systems allow. Daisy-chaining stations, even on today's analog systems, is not the best approach anyway because it can add noise and make failures harder to diagnose and fix. Although daisy-chaining is convenient when running cable around a large event, the main reason the tradition developed was the small number of output channels that intercom systems offered in the early days. Today's newer intercoms effectively eliminate the need for daisy-chaining because most of their digital switching matrixes can deliver 128 × 128 crosspoints and are often expandable to 512 × 512 or larger.

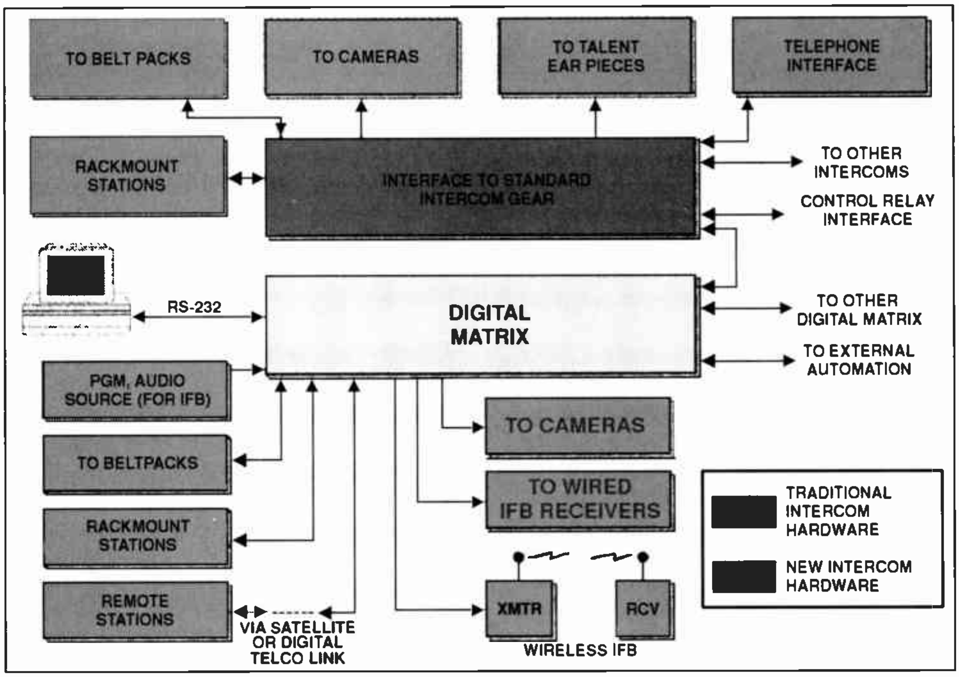


Figure 5.11-5. Block diagram of a fully featured current intercom system incorporating traditional equipment and new digital systems. Note the analog interface provided to adapt to existing hardware. (Courtesy *Broadcast Engineering*)

As digital hardware costs continue to fall, expect the trend to digital audio distribution to continue, especially in the case of large central crosspoint systems. The major advantage of taking digital audio all the way through the system is the elimination of induced electrical hum and noise. This can be appreciated by anyone who has ever had to run intercom lines around a pile of lighting cables and SCR dimmers.

Finally, digital does not necessarily mean serial digital (coax) installation. Depending on the application, being able to interconnect the stations with runs of coax may offer a lot of advantages. Check with the manufacturer regarding their definition of digital.

System Administrator

Intercom systems are supposed to just work, like an elevator, or the water faucet in the kitchen. They are supporting structures that allow the real work to happen more efficiently. As important as intercoms are to most people, it would be short sighted to minimize the attention that is required for them to become something that just works. When discussing the station's needs with manufacturers, identifying the stations engineering manpower resources do not specify a system that is so sophisticated or complicated that it becomes a maintenance or administration burden.

If the station's needs are very sophisticated, especially if there is computer software programming involved, consider assigning one person to the task of administration. It is almost impossible to maintain a large or complex system if more than one person can make fundamental changes to the software. This person typically need not be electronics expert, but rather more like a computer network manager. Maintaining proper filenames, backup files, entering the changes in a timely manner, and communicating in written form the profile of the system to the users are the most important tasks for the administrator. If the administrator is not a member of the regular engineering staff, be sure he works closely with the engineering department so there is some personnel redundancy.

Another point to remember is that unlike hardware based systems, a software based centralized architecture intercom depends on the integrity of the software that defines its nature. Do whatever it takes to assure that the software is what it needs to be.

Cost-basis Analysis

The cost structure of the two systems, distributed crosspoint and centralized crosspoint, typically means that a centralized crosspoint architecture is more appropriate for a large facility, while the distributed architecture approach may be more cost effective for a smaller facility. The single most expensive component in the centralized architecture approach is the frame containing the crosspoint circuitry. In this design, the individual stations are typically half to one third the price of an equivalently sophisticated distributed architecture user station. So, if the system needs to be big, it gets very expensive to implement a distributed architecture system. Conversely, If the system is small, it may be difficult to amortize the expense of the frame across only a few stations.

For example, a facility that buys 50 distributed crosspoint stations costing $6,000 each will spend $300,000 and have limited circuits. That same facility using a centralized architecture approach will spend $100,000 on a frame and $100,000 on stations at $2,000 each for the same total of $200,000. In this example, for less money the facility gets additional versatility and an expansion path that is only $2,000 per station.

Obviously, not all systems are this big and the numbers can greatly vary. It would be just as unwise to spend $50,000 on a frame to connect eight $2,000 stations together. In this case, it would be better to spend $48,000 on eight $6,000 stations.

These two architectures are not the only choices. There are some that combine features of each, however, the industry trend is to focus on either great versatility and with it low fault tolerance and focused attention, or a limited, highly tolerant more consistent system.

Consider the Supplier

There are some manufactures that only produce one type of architecture while others make all types. Most large intercom systems, like those in a full size production truck or a large studio/news control room, use both distributed and centralized crosspoint systems. A large centralized crosspoint frame with 2-wire beltpacks interfaced with 2-wire to 4-wire converters is a common application over all systems. In addition, it is not uncommon for the 2-wire and 4-wire components to come from different manufactures because each is considered better at one technology than the other for a specified system.

Do not be intimidated by the complexity and versatility offered in today's intercom systems. One important goal may be to simply emulate the 2-wire system that the station's operators are most familiar with. Build from this common point, adding the features and future options later.

Intercom systems do not get much notice—until they fail. So, when all is said and done, if the new intercom system meets the operator's needs and does so reliably, it has been a successful endeavor.

KEY TERMS

Administrator. The person that maintains the functional integrity of a system.

Call. Feature allowing user to get the attention of someone at a remote or main station, typically through a light or tone.

Conference line. Also known as party line, everyone hears all communication (see 2-wire system).

Control data. An electronic information signal that allows an action at one location to effect the function of a device at another location. Some systems embed the control data into a digital audio data stream form one point to another.

Crosspoint. A switch in an audio signal path from one point to another that is opened or closed at the request of the user.

Crosspoint control. The method by which the user requests a crosspoint be opened or closed.

Channel. Most commonly refers to a 2-wire path with multiple users on the same wire.

Dry pair. A telephone line or any other conductive cable of two wires that can carry audio but no voltage.

Duplex. The ability to both talk and listen at the same time.

Two-wire (2-wire). A user station connection scheme where by the talk and listen for a particular user are on the same wire pair. Each pair of wires can consist of ground/common and a single conductor or in a balanced configuration, a positive and a negative conductor. Usually found in a distributed crosspoint architecture

Four-wire (4-wire). A user station connection scheme where by the talk and listen audio for a particular user are on two different wires. Usually a balanced audio pair is used for the talk out and a separate pair for the listen input. Signals can be carried in an unbalanced wire set while still being referred to as 4-wire. This technique is most commonly found in centralized crosspoint architectures.

ISO. Also called isolated feed. Typically a private feed to an important station on the intercom system.

Listen signal. The signal that emanates from a speaker, earphone or headset earpiece.

Matrix. A device that switches inputs to various outputs or channels. A matrix intercom is usually point-to-point, 4-wire product (see 4-wire system).

PL. Short for party line. Any one channel having multiple stations connected to it.

Point-to-Point. Often refers to a type of intercom system, as opposed to a party line type of intercom. Point-to-point communication occurs between one person/station to a single other person/station (see 4-wire system and matrix).

Port. An audio input and output pair with dedicated real world connections. Works with a 4-wire scheme.

Program input. An audio source, typically on-air or studio audio, fed to an input channel on the intercom station for monitoring by stations.

Side tone. The audio heard in a headset when a user speaks-their own voice.

Station. A device, either rack mount or portable, consisting at a minimum of a microphone connector/ preamplifier with an off/on switch, and a speaker with an amplifier/volume control. May also include serial data stream, and multiple listen/talk circuits.

Talk signal. The post-microphone preamplified signal that will be heard by other people.

2-wire system. A communications system in which the talk and listen signals to and from each station travel on the same path. Also called party line, PL, conference line or interphone.

4-wire system. A communications system where the talk and listen signals to and from each station travel on different paths (wires). Also known as matrix, squawk, point-to-point, talkback or intercom.

User. A person that talks or listens at their discretion to other individuals utilizing the intercom system.

Wet pair. A telephone line or any other conductive cable of 2 wires that carries DC voltage to power remote stations.

BIBLIOGRAPHY

McHaddad, Andrew, "Choosing an Intercom System," *Broadcast Engineering,* July, 1997, pp. 46-103.

Tourkow, Bob, "Intercom System Design," *Broadcast Engineering,* March, 1986, pp. 74-96.

Lies, Bennett, "Intercoms: No Longer an Afterthought," *Broadcast Engineering,* May, 1995, pp. 54-64.

Brand, Dave, "Basics of Intercom System Design," AES Seminar, November, 3, 1988.

Clear-Com Systems, "Party Line Intercom System Installation Manual," September, 1997.

Acknowledgements: The author wishes to thank Ed Fitzgerald, Gary Parks and Bob Tourkow from Clear-Com, John King from RTS/Telex as well as Bennett Liles and Andrew McHaddad for their assistance with this work.

5.12
WEATHER RADAR SYSTEMS

ROBERT BEACH
KAVOURAS, INC., BURNSVILLE, MN

INTRODUCTION

Radar was invented in the 1930s and has been widely used for air traffic surveillance and control. Beginning in the late 1940s, it was determined that radar could also be useful for detecting and measuring precipitation. Since then radar has become the preferred technology for detecting and tracking severe weather. More recently, Doppler capability has been added to many weather radar systems to permit measuring the wind velocity associated with the precipitation.

At this writing, the U.S. National Weather Service (NWS) was in the final stages of upgrading the U.S. National Weather Radar network to WSR-88D (NEXRAD) radar systems. WSR-88D is a sophisticated fully coherent Doppler weather radar with true 3D volume scanning and processing capability.

Also at this writing, the U.S. Federal Aviation Administration (FAA) was installing Terminal Doppler Weather Radar (TDWR) systems at selected airports in the U.S. TDWR is also a sophisticated, fully coherent Doppler radar system whose primary function is to detect low level wind shear in the approach and departure flight path of an airport runway. Low level wind shear has proven to be extremely hazardous to aircraft during takeoff and landing.

BASIC RADAR THEORY

The term *radar* is an acronym for *RA*dio *D*etection *A*nd *R*anging. There is no magic; radar is nothing more than a very special way of using radio signals.

The basic principle of radar is simple; transmit a pulse of radio signal, form that pulse into a very precisely focused beam and listen for reflections that bounce back from any targets the beam happens to come in contact with. For a weather radar, the desired target is precipitation: rain drops, hail stones, snow flakes, etc. The larger and/or more numerous the precipitation particles, the stronger the reflection.

The reflections are typically displayed on a plan position indicator (PPI) display as colored areas where the color represents the strength of the reflected signal. The PPI display is essentially a map with the radar at its center. Thus, intensity of precipitation, in a given geographical location within the radar's range, can be estimated quickly.

A radar with Doppler capability also measures the speed at which the precipitation is moving. Moving precipitation implies wind, thus a Doppler radar can also estimate wind velocities. In a PPI Doppler velocity display, the wind velocity is displayed with the color indicating the wind velocity in each geographical location.

In addition to forming the transmitted signal into a very precisely focused beam, the radar antenna also aims that beam into the area of interest in the atmosphere. Typically, radar antennas will be set to scan the atmosphere at the rate of 2 or 3 revolutions per minute. If the radar beam were visible, it would look very much like the beam of light from a lighthouse sweeping around through the atmosphere.

Wavelength

There is a very narrow range of radio wavelengths that are useful for weather radar. If the wavelength is too long, there are no useful reflections from precipitation. If the wavelength is too short, the signal is absorbed rather than reflected by intense precipitation. The range of useful wavelengths is between 3 and 10 cm.

Table 5.12-1
Wavelength frequencies used for weather radar.

Wavelength Frequency	Band	Typical Radar	Comments
10 cm (3 GHz)	S	National Weather Service WSR-88D (NEXRAD)	Good penetration of torrential rainstorms, but requires very high power and a very large antenna.
5 cm (6 GHz)	C	FAA TDWR, and most commercially available weather radars	Reasonable trade-off of operating characteristics, works well with modest antenna size.
3 cm (10 GHz)	X	On-board aircraft weather radar systems	Small antenna compatible with aircraft space limitations, but absorption of signal by intense precipitation can produce misleading images.

Locating the Reflection

Radar systems locate the target with respect to the radar antenna (the direction and distance from the radar antenna to the target). The distance is typically referred to as the *range*.

Direction

The radar signal processor monitors the azimuth position of the antenna as the antenna rotates. Thus, the direction of transmitted and received pulses, and consequently the direction to each target is known.

Range

The range (distance) from the radar antenna to the target is determined by measuring the length of time that elapses between transmitting the pulse and receiving the reflection. Because the propagation velocity of the signal is a known constant, the distance from the radar antenna to the target can be accurately calculated.

Many radar systems use *range bins* to determine the range of target reflections. The radar's range is divided into some number (typically somewhere between 100 and 1024) of *bins* of equal size. Each bin has a time associated with it that corresponds to a distance from the radar antenna. The receiver's signal is sampled at each range bin time, and the strength of the signal at that time is stored in that bin. The resulting array of data represents the target reflectivity versus distance along the radar beam.

Pulse Rate and Range

All the reflections of interest should get back to the radar before the next pulse is transmitted. Therefore the pulse rate and the radar's maximum range are related. The pulse rate, in pulses per second, is often referred to as the PRF (pulse repetition frequency).

That relationship is expressed by the following formula:

$$R_{max} = \frac{c}{2(\text{PRF})} \tag{1}$$

Where:

R_{max} = Maximum unambiguous range
c = Speed of light (typically 300,000 km/sec)
PRF = Pulse repetition frequency (pulses per second)

For example, the maximum unambiguous range at a PRF of 1,000 pulses/sec is:

$$R_{max} = \frac{300{,}000}{2(1{,}000)} = 150\text{ km}$$

From this relationship it is clear that longer ranges require lower PRF's.

If a reflection from a previous pulse returns to the radar after the next pulse has been transmitted, it is frequently referred to as a *second trip echo* and the range at which it is displayed is said to be *folded*. The term *unambiguous range* means the range at which all reflections are first trip echoes with no range folding. In the above example, with the radar set to a pulse rate for 150 km range, a target located at 200 km range producing a very strong reflection might appear to be at 50 km range. Range folding can be confirmed by setting the radar to its longest range and observing the data for one or more scans. Targets producing strong reflections which appear to move out to the longer range are almost certainly folded when they appear at shorter ranges.

Reflection Strength

The strength of the reflected signal is proportional to the number and size of precipitation particles within the radar beam, and to the type of precipitation: rain, snow, hail, etc. The reflection strength is also proportional to certain parameters in the radar system: transmitter power, pulse width, etc.

Typically, the weather radar's signal processor categorizes the strength of the reflected signal into one of 16 levels. Each level is displayed in a different color. Therefore, it is easy to look at a display and quickly determine the strength of the reflection in any given location. Typically the weakest reflections are shown in light blue, with a spectrum of colors up to bright red for the most intense reflections. For meteorological purposes, these levels are usually measured in units of "Z". Z is the radar reflectivity factor for precipitation.

Z is usually displayed as decibels of Z. This is done because going from the weakest to the strongest Z routinely covers six or seven orders of magnitude of signal strength. The decibel scale provides a convenient way of expressing such a wide variation. The standard reference for Z is the signal reflected by precipitation with a single 1 mm drop of rain per cubic meter of space.

A signal strength of +10 dBZ means the reflected signal is 10 times stronger than the signal reflected by the 1 mm drop/m^3 standard. +20 dBZ is 100 times stronger, etc. A −10 dBZ signal is 10 times weaker than the standard, etc. A 0 dBZ signal is equal to the standard. Figure 5.12-1 provides a comparison of the sensitivity versus range of a typical commercial weather radar with the sensitivity of both the TDWR and WSR-88D radars.

The actual amount of power reflected back from precipitation to the radar can be calculated using the *Probert-Jones* equation. In practice, the radar set measures the power reflected back (P_r, the received signal) and solves the Probert-Jones equation for the Z value that caused that reflection. The precipitation intensity is then estimated from that Z value.

$$P_r = \frac{P_t G^2 \theta^2 H \pi^3 K^2 L Z}{1024(\ln 2)\lambda^2 R^2} \tag{2}$$

Where:

- P_r Reflected signal power received at the radar antenna
- P_t Transmitted power—fixed by transmitter design
- G Antenna gain—fixed by antenna design
- θ Antenna beamwidth—fixed by antenna design
- H Pulse width—limited to certain values by transmitter design
- K^2 Physical constant—typically 0.93 for rain or 0.197 for ice
- L Loss factors—determined by radar installation and atmospheric factors
- Z Target reflectivity—determined by the type and intensity of precipitation

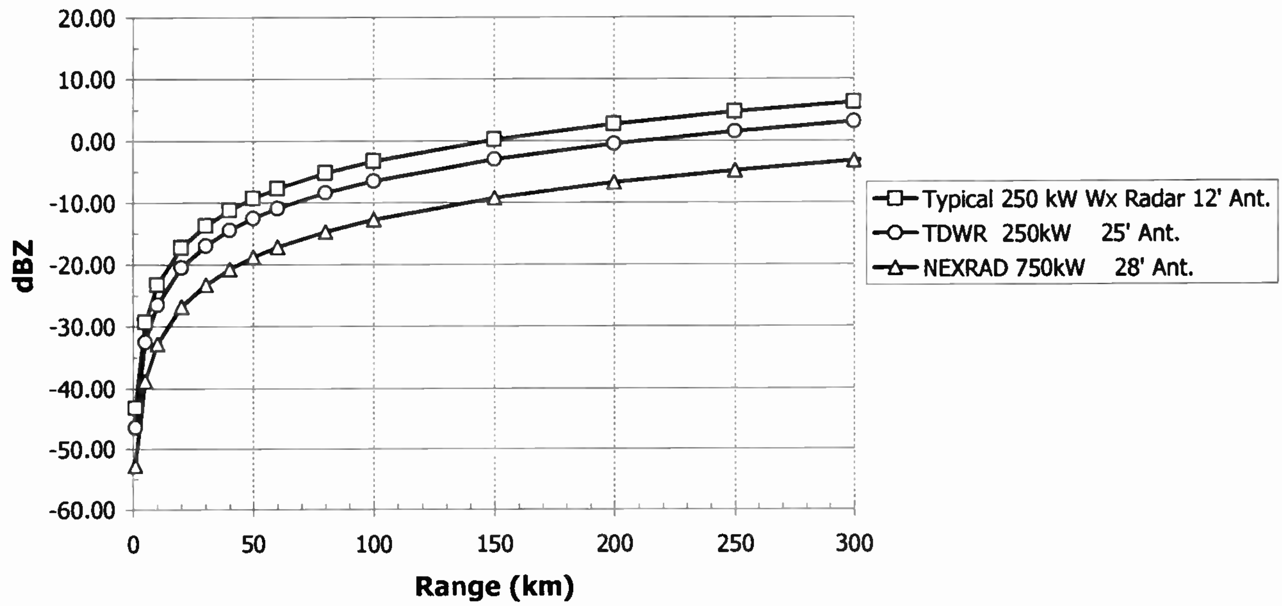


Figure 5.12-1. Sensitivity versus range.

λ Transmitter wavelength—fixed by transmitter design
R Range—distance from radar antenna to target

To be strictly accurate in calculating the reflected power, rain drops would have to be perfectly spherical. Since rain drops are not perfectly spherical, there is a small difference between the theoretical and actual results. This difference is accounted for in the empirical nature of the value of Z.

Pulse Width and Resolution

Weather radar systems typically transmit pulses with widths in the range of 1 μsec to 5 μsec. Some systems are capable of narrower and/or wider pulses.

The advantage of a wider pulse is that more signal power is reflected from the target, therefore it is easier to see weak targets with wider pulses. Doubling the pulse width is equivalent to doubling the transmitter's peak power. This is expressed by the "H" term in the numerator of the Probert-Jones equation.

The advantage of a narrower pulse is the ability to see the separation between closely spaced targets. To be separately distinguishable, targets must be separated by more than 1/2 the pulse length. As an example, the propagation velocity of a radio signal is approximately 1,000 ft/μsec. 1 μsec pulse will be approximately 1,000 ft from front edge to the back edge in the direction of its travel. Therefore, with a 1 μsec pulse, multiple targets that are separated by less than approximately 500 ft will appear on the radar display as one single target.

For many meteorological situations, targets of interest are frequently several miles wide and several miles apart. In this situation, the ability to use wider pulses to see weak targets may be more important than the ability to distinguish between several closely spaced targets.

Doppler Effect

The *Doppler effect* is named after Christian Johann Doppler, an Austrian physicist. Doppler became interested in why the sound of a train whistle seemed to change to a lower pitch as the train passed by where he was standing. He determined that, by measuring the change in pitch, it would be possible to accurately calculate the speed of the train. This principle applies to radio waves as well as sound waves. One of the more familiar examples of the Doppler effect is a police radar used to measure the speed of vehicles.

A Doppler weather radar measures the speed of the target as well as the target's location and reflectivity. The speed of the target (precipitation) implies a wind velocity.

The wavelength of the reflected radar signal is slightly compressed ahead of a target or slightly stretched behind a target, that is moving. The radar will measure either a slightly shorter or slightly longer wavelength reflection, depending on whether the target (precipitation) is moving towards or away from the radar antenna. After the reflection is received, the radar's signal processor compares the wavelength of the reflected signal to the wavelength of the transmitted signal. The velocity of the target is calculated from the difference in wavelengths.

Because the Doppler wavelength shift is very small, obtaining accurate Doppler velocity data requires a radar system with a high degree of frequency and phase stability. If the frequency or phase of the transmitted signal changes, even a little bit, during a pulse, there is no way for the signal processor to determine if the

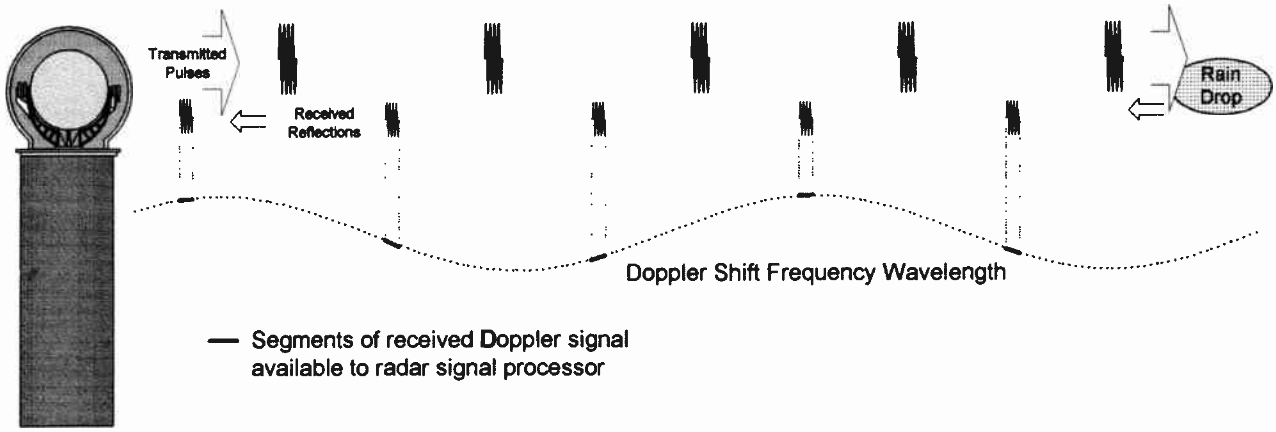


Figure 5.12-2. Doppler shift sampling technique.

difference it is measuring is due to real velocity of the target or to changes in the transmitted pulse.

Figure 5.12-2 illustrates the manner in which the radar signal pulses, in effect, sample the Doppler shift wavelength. The Nyquist data sampling principle requires that there must be a minimum of two samples per wavelength for the samples to accurately measure the wavelength. As a result, the maximum Doppler velocity any pulsed radar can measure is a function of the PRF (Nyquist principle) and the wavelength (Doppler principle). This is expressed by the following formula:

$$V_{max} = \frac{PRF\lambda}{4} \qquad (3)$$

Where:

V_{max} = maximum unambiguous velocity
PRF = pulse repetition frequency (pulses per second)
λ = transmitted wavelength (0.0535 m for C-band, 5600 MHz)

For example, for a C-band radar with a PRF of 1000, V_{max} is:

$$V_{max} = \frac{(1000)(0.0535)}{4} = 13.375 \text{ m/sec} \left(\approx 30 \frac{\text{mi}}{\text{hr}}\right)$$

Many radar systems provide velocity unfolding capability. These techniques often use staggered pulse rates and a processing algorithm which can multiply the maximum velocity that can be measured by a factor of 2. Some radars also offer a 3 times velocity unfolding capability.

From this relationship, it is evident that higher velocities require high pulse rates.

Maximum Unambiguous Range and Maximum Unambiguous Velocity

In a pulsed Doppler radar, the PRF, the maximum unambiguous velocity and maximum unambiguous range are interrelated. The graph in Figure 5.12-3 expresses that relationship for a C-band radar. Measuring high velocities requires a high pulse rate. Measuring long ranges requires a low pulse rate. Therefore there is a trade-off between the maximum velocity and the maximum range. This situation is sometimes referred to as the Doppler dilemma. The end result is that, at any given range, there is one unique PRF that will provide the maximum velocity measurement capability at that range.

Coherence

To obtain Doppler velocity data, the frequency and phase of the reflected pulses that are received must be compared to the frequency and phase of the transmitted pulse.

There are two challenges in making this comparison:

- The Doppler frequency shift is very small, so this comparison requires a high degree of stability and accuracy to get reliable velocity data. A great deal of precision is required in the radar's transmitter and signal processing circuits.
- The received pulse arrives back at the radar long after the transmitted pulse has ended. Therefore, it is impossible to make a direct comparison between the transmitted and received pulses.

There are two commonly used ways of making an indirect comparison between the transmitted and received pulses:

- A *coherent-on-receive* radar typically uses a *magnetron tube* in the transmitter. The magnetron is a power oscillator which starts up, transmits a pulse and then shuts down. This cycle is repeated for each pulse. Thus each pulse can have a slightly different frequency, depending on magnetron voltage, current, temperature, magnetic field, etc. The phase of each pulse from a magnetron is random.

  Circuits, typically called stable local oscillator (STALO) and coherent oscillator (COHO), sample and remember the frequency and phase of the magnetron's last pulse. The output of the COHO is used

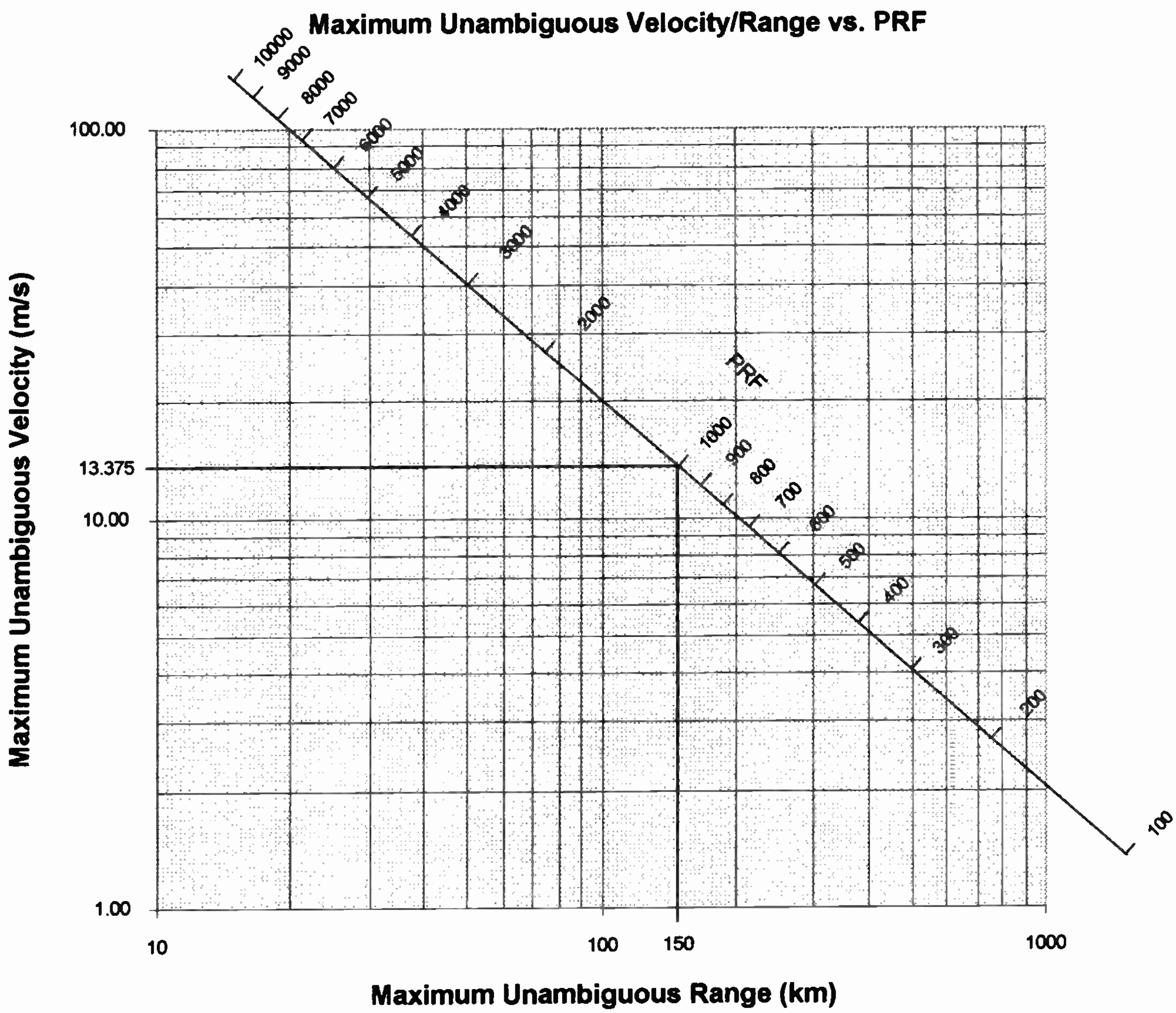


Figure 5.12-3. C-Band Doppler weather radar PRF vs. range and velocity.

as the reference for Doppler processing of the received pulses.

Many commercially available weather radar systems are coherent-on-receive. As a result of the small variations between pulses, these systems are typically limited to around 30 dB of ground clutter suppression by clutter filters.

- A *fully coherent* radar has a precision oscillator which operates continuously. This oscillator serves as the source of a master reference signal for all circuits in the radar.

The transmitter is a true amplifier which typically uses a klystron tube, traveling wave tube (TWT) or solid-state device(s). Each transmitted pulse is a small portion of the reference signal that is amplified, thus the frequency and phase of transmitted pulses are accurately known and are consistent from pulse to pulse.

The signal processor uses the precision oscillator as the reference for determining Doppler shift in the received pulses. As result of the phase locked consistency of all signals, fully coherent radars can typically achieve around 50 dB of ground clutter suppression from clutter filters.

WSR-88D, TDWR, nearly all research radars and a few commercially available weather radar systems are fully coherent.

BASIC RADAR METEOROLOGY

Radar meteorology is a science in itself. The information presented here contains only the most basic concepts. The bibliography at the end of the chapter provides references to far more detailed and comprehensive information.

Reflectivity and Precipitation Intensity

Interpreting the meaning of radar reflectivity displays requires meteorological knowledge plus experience with local conditions. As an extreme example of local conditions, there are places where, at sunset, large numbers of bats leaving their daytime roosts may appear on the radar as a strong gust front line.

The amount of signal reflected is determined by the number, size and type (rain, snow, hail, ice crystals, etc.) of precipitation. Light snow is a very weak reflector of radar signals, hailstones that are wet on the outer surface are very strong reflectors, and liquid raindrops are somewhere in between depending on the size of the drops. Research has shown that precipitation reflects horizontally polarized signals best. Nearly all weather radar systems in use today have antennas which transmit horizontally polarized signals.

With current technology, the radar can only measure the total reflection. Thus estimating the actual amount of precipitation from only a radar image becomes a classic case of one equation with three unknowns.

There is a significant amount of research currently being done to determine how get more precipitation information using a radar system. Most of that research is focused on radar with dual polarization capability, that is a radar that can transmit pulses that alternate between horizontal and vertical polarization. The differences between the horizontal and vertically polarized reflections are analyzed to extract more information about the precipitation. Differential reflectivity (ZDR) has been demonstrated to have a high probability of being able to accurately differentiate between hail and other types of precipitation. Research suggests that ZDR, in conjunction with linear depolarization ratio (LDR), differential phase shift (KDP) and other parameters will ultimately be able to provide much improved precipitation data.

The reflectivity of raindrops is approximately proportional to the diameter of the drop raised to the 6th power, while the volume of water in the drop is approximately proportional to the diameter of the drop raised to the 3rd power. Thus the relationship between reflectivity and actual water content of a raindrop is highly non-linear. The reflectivity is also proportional to the number of drops; twice as many drops reflect twice as much signal. Since neither the drop size nor the number drops can be readily determined, calculating the actual water content of precipitation from radar reflectivity data alone is at best an estimate.

One commonly used method of estimating rainfall rate from radar images is to use the formula:

$$Z = aR^b \qquad (4)$$

Where:

Z = radar reflectivity factor of the precipitation
R = rain rate in mm/hr
a = an empirically determined constant
b = an empirically determined constant

It should be emphasized that this formula provides only an estimate of precipitation. Over the years there have been a large number of experimental data sets collected to attempt to refine the values of a and b in this relationship. Table 5.12-2 provides rainfall rates in inches/hour for various commonly used Z-R relationships:

Table 5.12-2

Radar rainfall precipitation estimates—inches/hour.

DBZ	$300R^{1.4}$ Stratiform and convective rainfall	$200R^{1.6}$ Stratiform rainfall	$486R^{1.37}$ Convective rainfall (reduces hail effect)
20	0.02	0.03	0.01
30	0.09	0.12	0.07
40	0.48	0.47	0.36
50	2.5	1.9	1.9
60	12.9	8.1	10.3
70	67.0	34.1	55.4

Many radar systems allow the operator to enter values for a and b and display reflectivity directly as estimated rain rates. Some radar systems also provide a capability to total the estimated rainfall over a period of time.

As a point of reference, for a number of years the NWS used the values given in Table 5.12-3 to correlate display color, estimated rain rate and reflectivity dBZ:

Table 5.12-3

Display system color definitions.

DVIP Level	Rain rate (in/hr)	Reflectivity (dBZ)
1- lt. green	0.1	0–30
2-dk. green	0.25	30–41
3-yellow	0.5	41–46
4-orange	1.25	46–50
5-red	2.5	50–57
6-dk. red	4.0	57 or more

It should be cautioned that these relationships only provide estimates of the water content based on radar reflectivity. In any given storm, the actual water content may vary significantly from the values indicated by the formulas and data shown here.

Reflection Shape

The shape of the area of precipitation is also significant. For example:

- If a line of precipitation begins to bulge out, the bulge (or bow echo) may indicate the presence of strong winds, possibly the outflow from a downburst
- A hook shaped echo may indicate the presence of a tornado
- A long thin line ahead of an area of precipitation may indicate a gust front.

Non-Meteorological Reflections

Weather radar systems from time to time will also display reflections from non-meteorological targets such as dust, insects, birds and even bat migrations. Correct interpretation of these factors is usually a matter of local experience.

Velocity

A Doppler radar measures the velocity of the precipitation's movement. That velocity implies a wind blowing the precipitation. Thus, a Doppler radar can measure wind velocity. The velocity the radar measures is either directly toward or directly away from the radar antenna. This is called *radial velocity*. For winds blowing at an angle to the radar beam, the velocity measured by the radar will be the vector component of velocity along the centerline of the radar beam.

In the wind velocity display, velocity is usually indicated by color. Typically winds blowing toward the radar antenna are shown in shades of blue and green while winds blowing away from the radar antenna are shown in shades of orange and red. Each color indicates a range of velocities, typically the darker/brighter the color, the higher the velocity.

Velocity measurement capability has dramatically improved the accuracy of tornado warnings. Research has shown that most tornadoes are produced by mesocyclones (rotating thunderstorms). The Doppler velocity measurements can usually identify the rotation within a storm cell many minutes before a tornado actually forms. This rotation is typically identified by winds within a storm cell blowing in opposite directions in close proximity to each other.

The velocity measurement capability can also detect winds associated with low level wind shear. Low level wind shear has been determined to be extremely hazardous to aircraft during takeoff and landing. Some Doppler radar systems can be programmed to automatically detect the patterns of storm related wind shear, and produce an alarm when those wind shear patterns are detected.

Spectrum Width

Spectrum width is the difference between the maximum and minimum wind velocities in close proximity to each other. If the wind were perfectly uniform, all velocity measurements in the same area would be identical and the spectrum width would be zero. In essence, spectrum width is a measure of turbulence in the atmosphere.

Many radar operators will use the spectrum width display to locate areas where development of severe weather is most likely. Areas of high spectrum width are also of interest to pilots, as areas of turbulence to be avoided.

Beam Width and Beam Altitude

There are several factors related to the antenna beam width and altitude:

- Radar theory requires that, for accurate measurement of precipitation intensity, the precipitation must completely fill the pulse volume of the radar beam, that is the height, width and length of the volume of space the transmitted pulse occupies. For many meteorological purposes, precipitation areas more than 3 km (1.86 mi) wide are of most interest. To assure beam filling, and therefore accurate measurement of precipitation, the radar beam should be no more than 3 km wide in the areas of greatest interest.
- As a result of the earth's curvature, the altitude of the radar beam increases with distance from the radar antenna. Most precipitation occurs at relatively low altitudes, typically below 10,000 ft, and nearly always below 20,000 ft. At very long distances, the radar beam is too high to provide any meaningful data.

  A useful rule of thumb is that, with the antenna elevation set at the normal $+0.5°$ tilt, the beam will rise approximately 1,000 ft for every 10 mi of range.

The diagram in Figure 5.12-4 indicates the beam width and altitude for a radar with 1° beam width (C-band with a 14 ft antenna at $+0.5°$ elevation angle). Figure 5.12-5 illustrates the altitude of the beam centerline as a function of range at various antenna elevation angles.

The diagram in Figure 5.12-6 illustrates two situations in which the radar can be misleading as a result of beam altitude. The first shows virga (streaks of water drops or ice particles which evaporate before reaching the ground) which the radar detects. However, no rain actually reaches the ground, even though it is displayed on the radar. The second situation, which is typical at longer ranges, shows the radar beam well above precipitation. Thus that precipitation will reach the ground, but will not appear on the radar.

Ground Clutter

Ground clutter is one of the largest single problems for a weather radar. It often masks real precipitation. For a radar covering a metropolitan area, the ground clutter is likely to be worst in the area where radar data is of most interest (directly over the metropolitan area).

Unfortunately, some of the transmitted pulse is radiated by the antenna at angles away from the main beam. These are called *side lobes*. Figure 5.12-7(a) illustrates how these side lobes cause reflections from buildings, trees, vehicles, etc.—these reflections are referred to as ground clutter. Figure 5.12-7(b) illustrates how reduced side lobe energy avoids much ground clutter. The antenna's first side lobe is usually the primary cause of ground clutter.

Stationary ground clutter from buildings, terrain, etc. can be filtered out either by using a *clutter map* to subtract known clutter sources from an image, or by using *clutter filters* in a Doppler radar. The effectiveness of Doppler ground clutter filtering is heavily dependent on the stability of the radar transmitter.

Filtering moving clutter caused by vehicles, sea surface waves, etc. is an extremely difficult problem. If moving clutter is expected to be a problem in a given location, selecting a radar antenna with the lowest possible side lobe energy is likely to be the most effective way to suppress that clutter.

Anomalous Propagation

Anomalous propagation (AP) is caused by atmospheric effects bending the radar beam downwards so

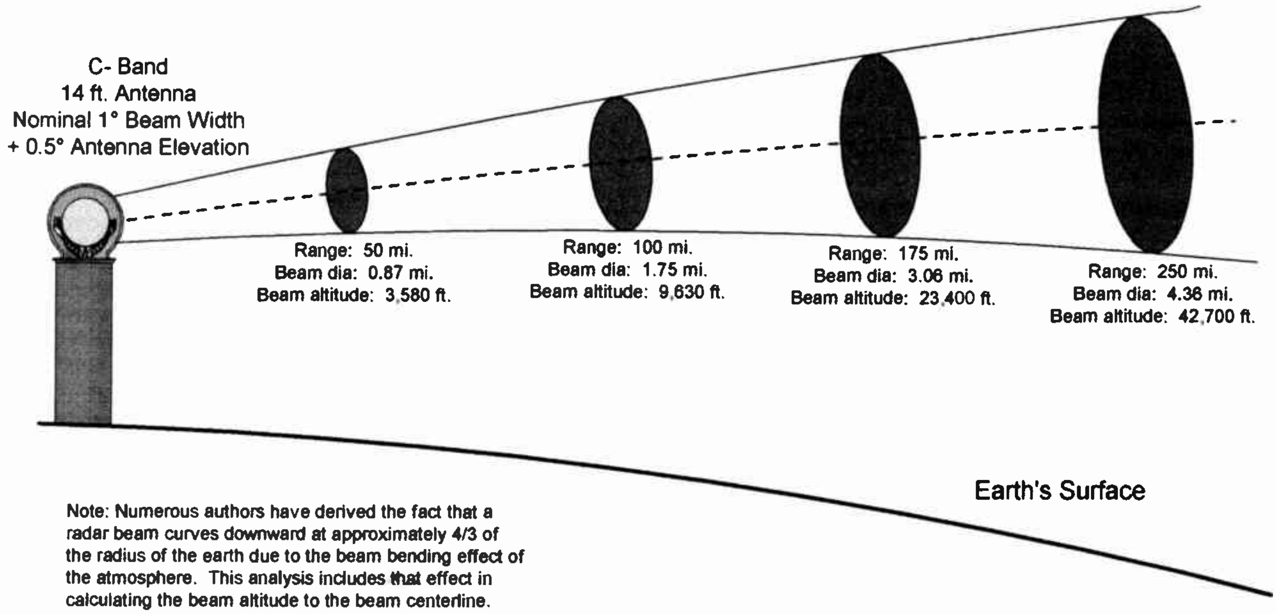


Figure 5.12-4. Radar beam size and altitude vs. range.

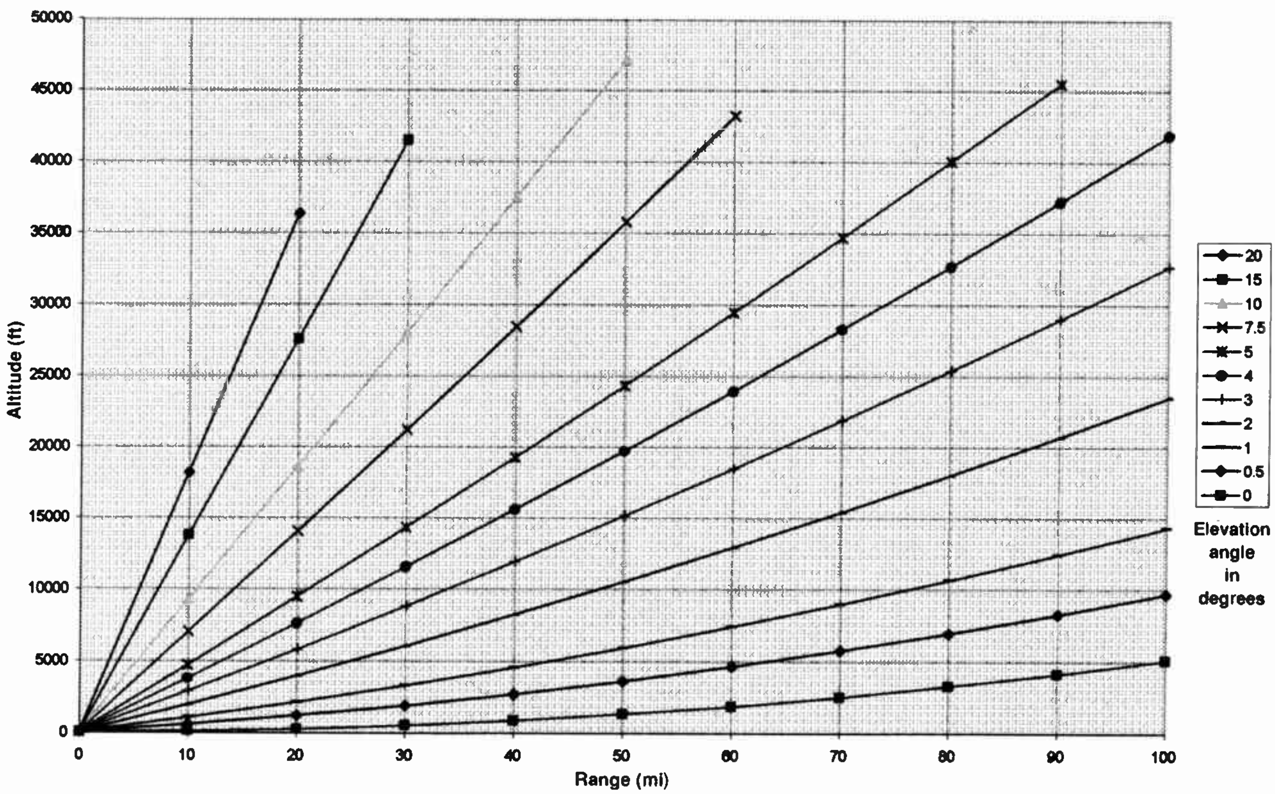


Figure 5.12-5. Beam altitude vs. range at various antenna elevation angles.

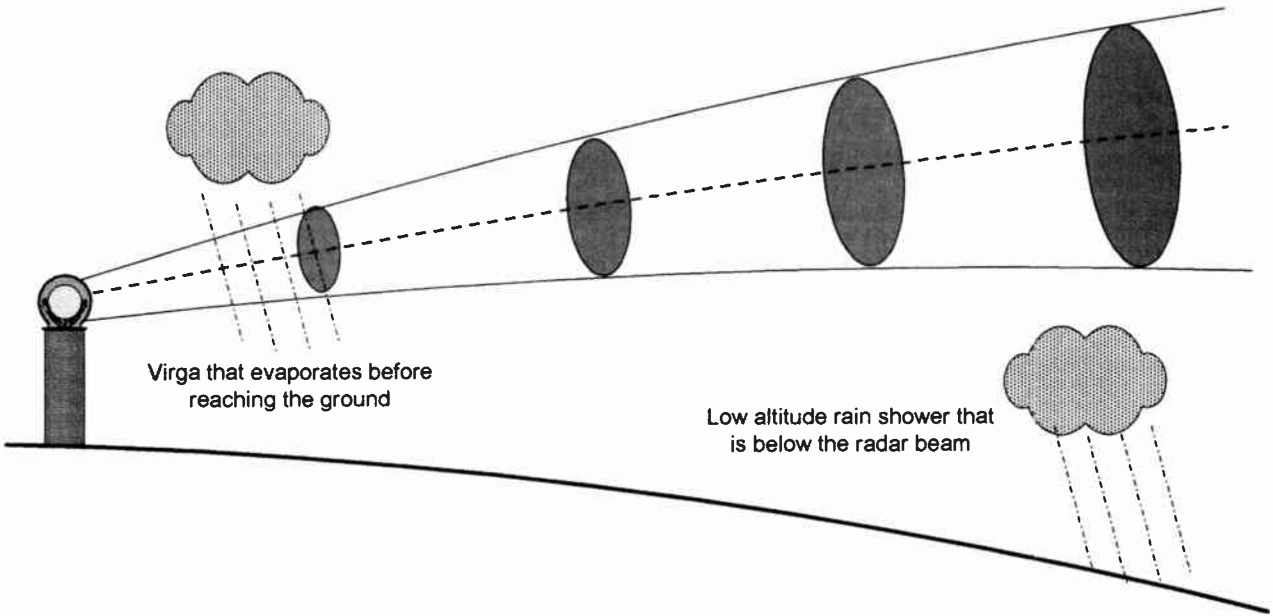


Figure 5.12-6. Rain detection considerations.

far that it actually hits the ground. The reflections travel back via the same path so that the image displayed is actually the ground. This is illustrated in Figure 5.12-8.

AP is typically a very low altitude situation caused by the radar beam passing through a thermal inversion layer. Raising the antenna elevation angle from the normal 0.5° to 1.0° or even 1.5° (getting the beam above the inversion layer) may help to identify AP. If raising the elevation angle causes a reflection area to move away from the radar or disappear entirely, that reflection is likely to be AP.

RADAR SYSTEM COMPONENTS

The *antenna* is a critical element in the overall performance of a weather radar system. The antenna performs two functions: forming the transmitted pulse into a very narrow beam and aiming that beam in the desired direction. If a weather radar beam were visible, it would look very much like the beam of light from a lighthouse, a very powerful, narrowly focused beam sweeping around through the atmosphere.

Weather radar antennas are almost universally parabolic. The one exception is small *slot array antennas,* sometimes referred to as pizza pan antennas. For most weather radar purposes, the performance of a properly designed slot array antenna is equivalent to a parabolic antenna of the same diameter.

For a weather radar, the general premise is that the narrower the beam the better. However, basic physics requires a progressively larger antenna to obtain a progressively narrower beam. The costs of a radar antenna, radome, tower, etc. increase dramatically for larger antenna sizes.

Figure 5.12-4 illustrates the relationship between range, beam width and beam altitude. From this data, it is apparent that a weather radar antenna with an angular beam width of approximately 1° will provide

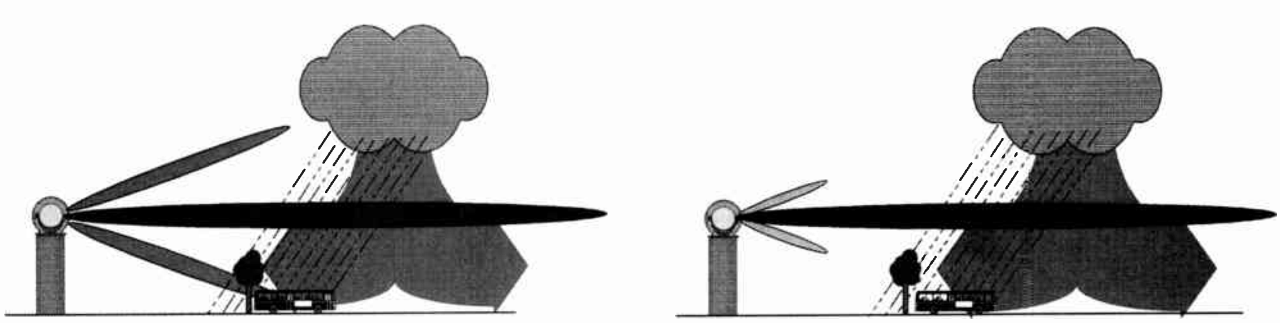


Figure 5.12-7. The effects of sidelobes on radar performance (a) large sidelobes, (b) small sidelobes.

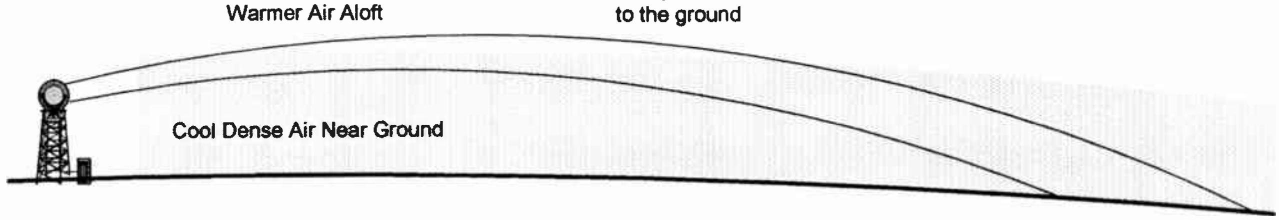


Figure 5.12-8. Radar beam super-refracted down to the ground.

a beam width of less than 3 km at ranges and altitudes that are often of the most meteorological value.

Side lobes are caused by a combination of basic physics, plus very small imperfections in the shape of the antenna, plus the waveguide, support struts and feedhorn being located in front of the antenna reflector where they are in the main beam.

Side lobe suppression is typically measured as decibels of side lobe signal strength, with the center of the main beam as the reference. This is typically quoted as the one-way value. The actual signal path is a round trip (transmitted signal out, and received signal in). The round trip attenuation is double the normally quoted value. Center feed parabolic antennas typically achieve side lobe suppression in the range of −25 dBc (25 dB down from the carrier) to −30 dBc one way. (The larger the negative number, the greater the attenuation.) Precision antennas with an offset feed arrangement have been reported to have side lobe suppression of −40 dBc or better one way.

Pedestal

The weather radar antenna is typically mounted on a device called a *pedestal*. The job of the pedestal is to accurately turn and elevate the antenna to aim the antenna's beam in the desired direction.

Typically the pedestal can turn the antenna continuously a full 360° either clockwise or counter-clockwise in the azimuth direction at a specified rate. The elevation (up and down) angle is often limited by drive mechanics to something like −2° to +60°, with horizontal being the 0° reference. This creates a situation where the area directly above the radar cannot be scanned. This area is sometimes referred to as the *cone of silence*. Since most weather occurs at lower altitudes, (at lower antenna elevation angles) the cone of silence is not a very significant limitation in most weather radar installations.

The pedestal sends both the azimuth and elevation angles to the radar processor. Frequently ±0.1° (1/10 of the antenna's beam width) is specified as the required accuracy for the antenna position and position measurement. Improving the pedestal's angular accuracy can be done, but the cost increases dramatically, and typically does not significantly improve the meteorological quality or usefulness of the data.

Most radar systems permit adjusting the speed at which the antenna turns. Turning the antenna faster provides more frequent updates for data at a given location. However, it also puts greater demands on the radar signal processor (one revolution's data must be processed in less time). If the antenna rate is too high, the accuracy of the Doppler velocity data is reduced. Typical scan rates are often in the range of 2 or 3 revolutions per minute. Many antennas are capable going as fast as 5 or 6 rpm.

Many older designs of radar antenna pedestals use dc motors, which require periodic replacement of motor brushes, to drive the antenna. Newer designs often use brushless motors. Because the pedestal is typically on top of a tower where it is not readily accessible, changing the brushes can be a significant undertaking.

Transmitter

A radar transmitter typically has three major functional areas: the high voltage power supply, modulator, and RF power output stage. In a Doppler radar, stability of the transmitter is particularly important because of the small changes that the Doppler shift produces in the reflected signal. If the transmitter causes small changes in the signal, it is impossible to separate those changes from the Doppler shift changes.

When evaluating radars, there is a tendency to place heavy emphasis on the peak power, with more being better. Peak power is only one factor. A radar having more a capable signal processor and an antenna with better side lobe attenuation may outperform the more powerful radar in many situations.

High Voltage Power Supply

The high voltage power supply typically produces a dc voltage somewhere between 10,000–75,000 V depending on the design of the particular radar. Peak pulse currents, at those voltages, may be 20 amps or more. While the peak power is very high, the average power is typically quite low, often not much over 100 W.

The high voltage must be well regulated and stable for the best accuracy of data. If the transmitted power, which is a function of the high voltage, varies the reflected power will also vary giving erroneous reflectivity readings. Variations in voltage may also affect the transmitted frequency and/or phase during a pulse, particularly with a magnetron transmitter. This variation in frequency will in turn appear as a Doppler shift in the reflected signal, giving erroneous Doppler velocity data.

Modulator

The job of the modulator is to form the high voltage into pulses of the proper duration and pulse repetition rate (PRF) to drive the RF power output stage. Many radars use a pulse forming network to do this. The pulse forming network typically uses some combination of capacitors and/or inductors to store energy such that a high energy pulse is available at the right time.

Pulse forming networks are typically designed around a particular pulse width and PRF. This factor limits a radar system that is based on pulse forming networks to a very narrow range of choices of pulse width and PRF.

Some research radars and a few commercially available radars use dc switch type modulators. A dc switch type modulator does not have the timing restrictions of a pulse forming network. Therefore a radar based on a dc switch type modulator typically allows the selection of pulse widths and PRFs over a wide range that is limited only by the duty cycle of the components involved.

Many older radars use Thyratron tubes to conduct the pulse from the pulse forming network to the RF power output stage. Newer designs often use solid state devices which tend to be much more reliable than Thyratrons.

RF Power Output Stage

There are two categories of RF power output stages: coherent on receive radars typically use magnetron tubes, and fully coherent radars which use a true amplifier that is often based on a klystron tube. Some commercially available fully coherent radars also use traveling wave tubes (TWT's) or solid state devices.

A magnetron can be described as a power oscillator. Every time a pulse arrives from the modulator, the magnetron starts up, begins to oscillate, produces a pulse at the RF frequency and power level and shuts down. This sequence is repeated for every pulse.

A magnetron operates by causing a beam of electrons to sweep around the interior of a cylindrical cavity. That cavity typically has a series of slots around the outer surface of the cylinder. The number of slots and the speed at which the beam sweeps around determine the operating frequency. The beam speed is determined by a combination of the physical geometry of the tube and slots, voltage, current, magnetic field strength and the load on the magnetron's output. As a result, the frequency of oscillation for the magnetron is likely to be slightly different for each pulse. The phase of each pulse is random.

Coaxial magnetrons have an additional cavity that couples the slot output to the waveguide output. This additional cavity tends to isolate the beam from output loads and make the operation of the tube more stable and less affected by changes in load.

A fully coherent radar has a master oscillator which is always running. The RF power output stage amplifies that signal during the pulse from the modulator. Thus the frequency and phase of each pulse is very stable, accurately known and essentially identical to every other pulse.

Klystron's and TWT's are both linear electron beam tubes. In both, the input signal is injected into the tube and the beam energy is used to amplify the input signal. The frequency and phase of the output signal is determined by input signal and is largely independent of the tube and circuit parameters in the RF power output stage.

Receiver

The reflections returning to the radar antenna are extremely weak. Therefore the receiver must be very sensitive and have a very low noise level such that weak signals can be detected and are not masked by noise. In addition, the receiver must be able to receive a very strong reflection from a nearby target without overloading. Meeting both objectives is a very challenging engineering problem.

Receiver sensitivity is typically referred to as minimum detectable signal (MDS) and is measured in dBm. Commercially available radar systems typically have an MDS in the range of −108 to −115 dBm (−115 is more sensitive than −108). A few systems have an MDS of −120 or better.

Dynamic range, the difference between the weakest and strongest signal the receiver can handle without overloading, is typically at least 90 dB. A few currently available systems can exceed 100 dB. In many systems, this dynamic range is achieved by using an intermediate frequency (IF) amplifier with logarithmic rather than linear response.

Unfortunately, the logarithmic amplifier alters the phase of the received signal, thus its output is not coherent and will not provide accurate Doppler data. Therefore, many Doppler radar systems include both a linear and a logarithmic IF amplifier. The log channel is used for reflectivity data and the linear channel is used by the Doppler signal processor.

Classical radar theory indicates that the IF bandwidth is related to the transmitted pulse width. Narrower pulses require wider IF bandwidths. Some radar systems automatically adjust the IF bandwidth when different pulse widths are selected.

Signal Processor

Most radar systems today use digital signal processors. The received signal is analog, so it must be converted to digital format by using an analog/digital (A/D) converter. In many radar systems the A/D converter has 12 bits of precision. The A/D process is critical in the overall performance of the radar system. The conversion process introduces noise and is subject to a number of factors which may reduce the quality of the signal.

The signal processor typically performs several functions:

- Performing $1/R^2$ compensation—Radio signals propagate according to the inverse square law, that is the signal has 1/4 the strength at twice the distance. If this were not compensated for, precipitation would always appear to be weaker at greater distances. The compensation causes precipitation of a given intensity to be displayed at the correct intensity regardless of its range
- Doppler velocity processing—The signal processor compares the frequency and phase of the signal in each range bin with the frequency and phase of the transmitted pulse and computes the radial velocity from that difference
- Spectrum width processing—The signal processor computes the difference between the maximum and minimum velocity within the same range bin
- Clutter filtering—In a Doppler radar, a ground clutter filter assumes that targets with zero or near-zero velocity are buildings, smokestacks, etc., and suppresses those reflections.

If the radar is to produce true real time images, the signal processor must have the capacity to process all the data immediately as the antenna scans. Processing capacities of 500 million instructions per second (MIPS) or more are readily available.

Radar Data Product Generator and Volume Scanning

The signal processing referred to in the previous section applies to data from the radar at a single antenna elevation angle. Many state-of-the-art radars have the ability to perform true 3-D volume scans. This is typically accomplished by taking successive scans of one revolution each with the antenna at successively higher elevation angles and repeating the sequence from the ground up.

In many radar systems the volume scan data is processed by a separate workstation. That workstation will generate many different radar image products from the volume scan data. A few typical volume scan products are:

- CAPPI—constant altitude plan position indicator, a horizontal slice through the atmosphere at an arbitrarily specified altitude
- VIL—vertically integrated liquid, the integrated value of the total amount of liquid in the atmosphere above a specified point on the ground
- XSECT—cross-section, a vertical slice through the atmosphere above a line between two arbitrarily located points on the ground
- VAD—velocity azimuth display, a display of a time series of wind velocity versus direction and altitude parameters.

A number of other volume scan products have been devised for various purposes. Radar manufacturers can provide details of the products their particular equipment supports.

Display

The basic images a Doppler weather radar system displays are reflectivity, velocity and spectrum width. Some systems provide both raw and filtered reflectivity. The operator controls which image is displayed.

The reflectivity, velocity and spectrum width displays are typically shown with specific colors representing calibrated data values. Thus precipitation intensity, wind velocity and atmospheric turbulence at a given location can be quickly estimated.

The basic radar display contains a great deal of information, but is typically not very useful by itself because the location of the data is not clear. Adding a map overlay to the radar image is one of the most common display graphic functions. A map permits users to quickly identify areas of precipitation, velocity, etc. with respect to known locations on the ground.

Many radars provide the ability to blink selected colors in a display. This makes it easy for users to identify areas of particular interest. Some radars provide the ability to generate additional graphic overlays that can be created by the user for specific purposes.

If a radar is to be used for live on-air presentations, it should have the ability to produce broadcast quality genlocked video output. For on-air presentations, having a radar system with a powerful graphics processor is a significant advantage. That graphics processor can be used for the creation of sophisticated weather products which have the look the station desires.

Uninterruptible Power System

Severe weather is one of the most frequent causes of power failures. Having an uninterruptible power system (UPS) for all radar equipment is good insurance against having the radar go dead.

A typical UPS system will have both a battery supply and an engine driven generator. The installation is typically designed such that the batteries will keep the minimum essential items running for a few minutes until the engine starts and can take the full load.

Some jurisdictions have special building codes that apply when large arrays of batteries are used inside of a building. The general concern is that, under certain conditions, most commonly available types of batteries produce hydrogen gas. In a small enclosed area, such as a radar transmitter building, that hydrogen can accumulate to an explosive concentration very quickly. If not covered by specific building codes, the UPS batteries should be in an area that is well ventilated, with that ventilation going directly to the outside and not connected to the rest of the building's ventilation system. In addition, the amount of electrical equipment in the battery room should be minimized and equipment in the battery room should be explosion proof.

LOCATION OF RADAR SYSTEM

There are many factors that affect the choice of the ideal location for a weather radar system and its antenna. However, the practical reality is, in many cases, only a certain location is available and that is where the radar must be placed. In these cases, effective use of the radar requires understanding the limitations imposed by the specified location.

Define Coverage

What is the primary objective of the weather radar?

- To provide severe weather warnings for a metropolitan area?
- To provide rainfall data for a particular watershed?
- To detect potential aviation hazards?
- To provide scientific data for a specific application?

Determining the nature of coverage to be provided and the primary area of interest for that coverage is required.

Selection Factors

If providing warnings of approaching severe weather is one of the objectives, more advance warning time is usually desirable. That dictates locating the radar system between the area to be protected and the usual direction of approach of severe weather. In many locations in the U.S., severe weather approaches from a generally westerly direction, therefore the radar in those locations should ideally be to the west of the area to be protected. Placing the radar elsewhere would typically reduce the warning time.

The radar beam from the antenna should have a clear view of the area to be protected. Locations where buildings, hills, etc., block or obstruct that view should be avoided. Trees also can block the view and trees grow. One common practice is to place the radar antenna on a tower that is tall enough so that it will still be comfortably above the tops of the trees after 20 years of growth.

EMC Considerations

The radar must be in an area that is free from interference on its operating frequency. Other radar systems are a common source of interference. Interference includes both other sources interfering with the radar's operation and the radar interfering with other equipment.

Non-RF sources can also be significant. For example, normal corona leakage from very high voltage power lines near the radar may raise the background noise level to the point that very weak radar reflections will be masked.

Site Access

The radar site location should have access 24 hours per day, 12 months per year. Selecting a site that is known to be periodically inaccessible as a result of snowfall, flooding, etc., is not recommended.

Radar to Control Location Data Link

In many situations the radar system must be placed in a location different from the location where the radar data is needed. In many broadcast situations, the radar system may be located at the transmitter site and the radar data and control must be at the studio.

Weather radar systems produce large volumes of data. Having real time coverage means getting that data from the radar to the studio in real time. That requires high bandwidth data circuits. It is not uncommon for a real time radar installation to require a T1 (1.544 Mbps) data circuit between the radar and the data/control location.

The availability and cost of high speed data circuits, or a private microwave or satellite link, between the radar site and the studio is a factor that should be considered in the site selection process.

Installation Planning

Planning the installation of a weather radar system involves a great many details. A checklist of items that may be helpful in the planning process is provided at the end of this chapter.

Lightning Protection

Lightning is one of the most frequent and costly causes of damage to weather radar systems. Systems from different manufacturers vary in their resistance to lightning strikes. However, nearly any radar installation will benefit from lightning protection which includes a grounding system suitable for the purpose.

Operating Frequency

An integral part of the plan should be the process of determining an operating frequency for the radar. That process is usually easier if an existing radar is being replaced. If the radar is new, a determination of a clear frequency must be made. Some radar manufacturers will require that the operating frequency be specified very early in the order process.

FCC

The radar system includes a transmitter which requires an FCC license to operate in the U.S. just as any other transmitter.

FAA

A tower for the radar antenna may be subject to FAA regulations, and require registration as an aviation hazard, based on its height.

Local Permits

Construction of a radar tower is likely to require local permits, variances and/or zoning changes.

SAFETY ISSUES

Radars are complex and powerful systems. Personnel who will operate and maintain a radar system should be trained in the safety aspects of the radar's operation as well as in the operational details of the radar.

Microwave Hazards

The effect of microwave radiation on the body has been the subject of a great deal of study. Of the potential hazards investigated to date, the heating effect on tissue is the only one that has produced general agreement and is readily replicated under controlled scientific conditions. The joint ANSI/IEEE *Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz* C95.1-1991 and the *FCC OET Bulletin 65* are based on tissue heating effects.

The location of the radar beam with respect to areas where people live and work must be considered. The radar antenna should be located such that no persons are within the area where the maximum permissible exposure (MPE) limits are exceeded. The MPE limit typically results in a minimum safe distance in front of the radar antenna given the transmitter power, antenna gain, frequency and duty cycle.

Electrical Hazards

The high voltage, high energy circuits in most radar transmitters present an extreme hazard to personnel. Because of the high voltage, normal skin resistance is ineffective at limiting current. Due to the circuits ability to deliver large quantities of energy very quickly, often in the sub-microsecond range, large amounts of energy will be conducted before even the fastest protective devices can operate.

Only personnel who are thoroughly trained should be permitted to work on a radar transmitter. Even then, strict safety procedures must be observed at all times. There may not be a second chance.

Mechanical Hazards

The motors that move the antenna are powerful and are typically geared down to provide the massive amounts of torque necessary to move a heavy antenna with precise accuracy. Typically the antenna is in a location out of sight of the circuits that control its movement. The antenna is also typically in a very confined space inside a radome. The end result is that personnel near the antenna have a high risk of being injured or crushed if the antenna moves unexpectedly.

Again, only personnel who are thoroughly trained should be permitted to work on the radar system, and strict safety procedures must be observed at all times.

X-Ray Hazards

Some radar transmitter tubes may produce X-rays in certain operating modes. The radar transmitter should not be operated if any X-ray shields provided by the manufacturer have been removed.

BIBLIOGRAPHY

Rinehart, R.E., *Radar For Meteorologists,* University of North Dakota, Knight Printing Co., Fargo, North Dakota, 1994.

Doviak, Richard J., and Dusan S. Zrnic., *Doppler Radar and Weather Observations,* Academic Press, Inc., Orlando, 1984.

ANSI, *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz.* Institute of Electrical and Electronics Engineers, Inc., New York, NY, 1994.

———; Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Fields (Bulletin 65), Federal Communications Commission Office of Engineering & Technology, 1997.

Weather Radar Site Checklist Site

1. **Area To Be Covered:** (Population centers, watershed area, airport, etc. prioritized in order of importance. A map with coverage areas outlined is helpful.)

2. **Type Of Coverage To Be Provided:** (Prioritized in order of importance)
   - General purpose meteorological data/forecasting
   - Severe weather detection/tracking
   - Rainfall measurement for hydrology application
   - Aviation safety hazard detection
   - Other ______________________________

3. **Antenna Site Location:** Lat. ______________ Long. ________ (Provide a brief description of location from known reference points if lat/long are not available)

4. **Radar Beam Obstructions:** (type, direction, approx. distance, consider coverage priorities)

5. **Access To Site:** (road, trail, cross-country, etc., special vehicles, equipment required? Is access subject to being cut off (snow, floods, etc.)?)

6. **EMC Factors:** (other radars, radio transmitters/receivers, etc. that may be interference problem and their location, also consider radar beam safety, i.e., personnel that may be exposed to beam during normal operation)

7. **Tower Height:** (radar must see over tops of trees and buildings, may be restricted by proximity to airport)

8. **Soil Conditions At Tower Site:** (sand, clay, rock, etc., soil test data available?)

9. **Electrical Service To Radar Site:**
   - Voltage, frequency, single or 3 phase: ______
   - Reliability of source: (voltage surges, sags, interruptions, etc., seasonal problems)
   - Number And Duration Of Interruptions In The Past 12 Months
     - Less than 1 second ________
     - 1 second to 5 minutes ________
     - More than 5 minutes ________

10. **Uninterruptible Power System (UPS):**
    - Battery Capacity: _____________ Watts/VA ______ Minutes
    - Generator Capacity: ___________ Watts/V/A ______ Hours fuel capacity
    - Testing/Maintenance Schedule Established:
    - Items powered by UPS
      - Radar Xmtr/Rcvr
      - Antenna Pedestal
      - Radar Data Comm.
      - Radar Workstation
      - Tower Warning Lights
      - Radar Air Conditioning
      - Security System
      - Other ___________________________

11. **Data Communications**

    Data communications capability will be required to get radar image data and control information back and forth between the radar and analysis locations.
    - Telcom service:

      Nearest connection point ______________________________

      Type of service ______________________
      (analog, ISDN, T1, E1, etc.)

      Data rate required/available ______________________________

      Interface required to connect to service ______________________________
    - Private microwave:

      Single hop or multi-hop ______________________________

      Obstructions in path ______________________________

      Multi-hop tower site locations ______________________________

      Rain fade potential ______________________________

      Radio interference ______________________________

      Data rates required/available ______________________________

      Interface required to connect to microwave ______________________________
    - Private fiber optic:

      Right-of-way for cable route ______________________________

      Interface required to connect to fiber optic ______________________________
    - Satellite data link:

      Antenna locations ______________________________

      Antenna look angles ______________________________

      Type of service required (SCPC, TDMA, etc.) ______________________________

      Data rates required/available ______________________________

      License requirements, restrictions ______________________________

      Rain fade potential ______________________________

12. Tower Grounding:
- Required ground resistance ______________ ohms
- Type of grounding system required:

13. Schedule:

Item	Date	Initials
Site location selected	_____	_____
Site location approved	_____	_____
Contractor selected	_____	_____
Drawings approved	_____	_____
Permits, licenses, etc. applied for	_____	_____
Permits, licenses, etc. obtained	_____	_____
Tower & building foundation done	_____	_____
Radar site electrical service in	_____	_____
Crane to erect tower	_____	_____
Pull cables	_____	_____
Install radar	_____	_____
Installation complete	_____	_____

5.13
CLOSED CAPTIONING AND EXTENDED SERVICES

AMNON M. SALOMON
GERALD FREDA
FALLS CHURCH, VA

INTRODUCTION

Closed captioning is the depiction of the audio portion of a television program as text displayed on a television screen with the aid of a decoder that may be internal or external to a television receiver. Closed, as opposed to open, captioning means that the captions do not normally appear as part of the broadcast television picture. The viewer must have the proper equipment and must select the captioning mode. Closed captioned programs are compatible with other programs in that the addition of the captioning signal does not interfere with the regular audio or video signals. Digital data to create captions are transmitted with the television program on line 21 in the vertical blanking interval (VBI).

Open captioning refers to captions that are transmitted as part of the picture rather than as a separate signal. Such captions are visible to all viewers (English subtitles in a foreign-language film.)

Captioned TV enables viewers to read the dialogue and narration of programs. The technique provides access to the entertainment, educational, and informational benefits of television for viewers who are hearing impaired. Other audiences can use captioning as well, including people learning English as a second language.

The captions produced by the closed captioning system generally appear in the lower portion of the television screen. The size of the characters varies in proportion to the size of the television screen. On a 19 in. screen, for example, they are about 1/2 in. high. The captions are easily visible—typically white letters against a black background—and usually do not obstruct essential parts of the picture (see Figure 5.13-1).

Closed captioning may be added in real time to a live program or it may be added later as part of post production or distribution. It may be done in-house or by a captioning service.

A Brief History of Closed Captioning

Closed captioned television technology was developed by the Public Broadcasting Service (PBS) during the period 1973–1979 with funding support from the Federal Government (Department of Health, Education and Welfare). Field test transmissions were conducted on all aspects of caption generation, encoding, decoding and display features of the service. PBS and others provided captioning services for several years after approval of the system by the FCC in 1976.

Figure 5.13-1. Digital closed caption data on video line 21 are processed by a decoder and displayed as text across the lower portion of the picture. (Courtesy of ABC News and the National Captioning Institute).

The closed captioning service was launched in March 1980 by the newly created National Captioning Institute (NCI), in cooperation with the ABC, NBC, and PBS networks, with 16 hours per week of captioned programming. The first consumer product containing the decoding feature, called *TeleCaption*®, was sold by Sears, Roebuck and Co.

The original target audience for captioning included hearing impaired people (about 20 million). The market has since expanded to include people learning English as a second language and those learning to read, especially students with reading disabilities.

FCC STANDARD

The line 21 captioning data signal is protected from interference from any other VBI service, test signal or spillover from active video under FCC Rules and Regulations, Section 73.682(a)(22), adopted in 1976. These rules also established the transmission standards for captioning and list the uses of the data channel.

Following are excerpts from rules associated with the Television Decoder Circuitry Act of 1990 that amended Part 15 of the FCC Rules (Radio Frequency Devices, which relate to television receivers).

Television Decoder Circuitry Act of 1990

The U.S. Congress passed the *Television Decoder Circuitry Act of 1990* (Pub. L. 101-431, 104 Stat. 960 (1990)). This act required that, effective July 1, 1993, all television receivers with picture screens 13 in. or greater must be equipped to display closed-captioned television transmission.

Excerpts from FCC Report and Order

General Docket 91-1*: In the Matter of Amendment of Part 15 of the Commission's rules to implement the Provisions of the Television Decoder Circuitry Act of 1990.* The Order was adopted April 12, 1991 and released to the public on April 15, 1991. The Order became effective July 1, 1993.

The Order amends Part 15 of the FCC's Rules. A new section 15.119 was added to read as follows:

Section 15.119 Closed Caption Decoder Requirements for Television Receivers.

(a) *Effective Date.* Effective July 1, 1993, all TV broadcast receivers with picture screens 13 inches or larger in diameter shipped in interstate commerce, manufactured, assembled, or imported from any foreign country into the United States shall comply with the provisions of this section.

(b) *Transmission Format.* Closed caption information is transmitted on line 21 of field 1 of the vertical blanking interval of television signals, in accordance with Section 73.682(a)(22) of this Chapter.

(c) *Operating Modes.* The television receiver will employ customer-selectable modes of *TV* and *Caption*. A third mode of operation, *Text*, may be included on an optional basis. The Caption and Text Modes may contain data in either of two operating channels, referred to in this document as C1 and C2. The television receiver must decode both C1 and C2 captioning, and must display the captioning for whichever channel the user selects. The TV Mode of operation allows the video to be viewed in its original form. The Caption and Text Modes define one or more areas (called "boxes") on the screen within which caption or text characters are displayed. . . .

(d) *Screen Format.* The display area for captioning and text shall fall within the SAFE TITLE AREA as defined by SMPTE Recommended Practice #27.3—1989. This display area will be further divided into 15 character rows of equal height and 32 columns of equal width, to provide accurate placement of text on the screen. Vertically, the display area begins on line 43 and is 195 lines high, ending on line 237 on an interlaced display. All captioning and text shall fall within these established columns and rows. The characters must be displayed clearly separated from the video over which they are placed. In addition, the user must have the capability to select a black background over which the captioned letters are displayed.

(l) *Compatibility with Cable Security Systems.* Certain cable television security techniques, such as signal encryption and copy protection, can alter the television signal so that some methods of finding line 21 will not work. In particular, counting of lines or timing from the start of the vertical blanking interval may cause problems. Caption decoding circuitry must function properly when receiving signals from any cable security system that was designed and marketed prior to April 5, 1991. Further information concerning such systems is available from the National Cable Television Association, Inc., Washington, DC, and from the Electronic Industries Association, Washington, DC.

(m) *Labeling and Consumer Information Requirements.* The box or other package in which the individual television receiver is to be marketed shall carry a statement in a prominent location, visible to the buyer before purchase, which reads as follows:

This television receiver provides display of television closed captioning in accordance with Section 15.119 of the FCC rules.

Receivers that do not support color attributes or text mode, as well as receivers that display only uppercase characters . . . must include with the statement, and in the owner's manual, language indicating that those features are not supported.

The complete text of the FCC *Report and Order* may be obtained from the U.S. Government Printing Office. The text was also published in the *Federal Register* Vol. 56, No. 114, p.27200. The FCC Rules & Regulations are contained in the *Code of Federal Regulations*, Part 47, Telecommunications. See Chapter 1.7, *Broadcast Standards* and *Recommended Practices*, for information on contacting the U.S. Government Printing Office and other organizations mentioned in this chapter.

Update Since Passage of Decoder Circuitry Act

Since the passage of the decoder circuitry act, several additions have been ruled by the FCC. These additions are documented with other technical details in a document titled *Recommended Practice for Line 21 Data Services* (EIA 608). This document can be obtained from the Electronic Industries Association, Engineering Department. A modest charge will be assessed for this document by Global Engineering Documents.

The regulation and use of line 21 for captioning has been expanded to include both fields. Additional features have been added to deal with additional languages. Within Field 2 is a capacity to deliver additional captioning services (C3, C4, T3, T4) in addition to extended data services (XDS). XDS includes but is not limited to content advisory (V-chip), Time of Day and source of transmission.

In June 1997, the next generation of closed captioning was adopted. This new standard (EIA708) will enable improved closed captioning to exist within digital television (including High Definition Television) while continuing to provide a method of delivering captions to analog sets via EIA608. The next generation caption system provides but is not limited to an improved appearance for captions, greater capability to provide multiple languages for a single program and greater user control (ability to select character size). The new standard was developed by an industry group made up of caption service providers, receiver manufacturers and encoder manufacturers.

Display Format

The line 21 captioning system has two independent channels, C1 and C2, each of which contains a captioning subchannel and a text subchannel. Caption data for most programming are carried in channel one, leaving channel two available for other applications such as caption data for another language. The captioning channel can operate in either of two different modes, *pop-on* and *roll-up*. Within each of these modes, *paint-on* can also be used.

The pop-on mode is used when the captions are prerecorded. Captions are loaded in advance of the time they are to appear and are displayed on command. While the caption in the first memory is displayed, a second memory may be loaded with the next caption for display upon the next command, and soon, alternately loading and displaying memories.

The roll-up mode is used for real time captioning. From one to four rows at the bottom of the screen display the real time captions with the text moving from bottom row upwards to the top-most of the 4 rows after which the text is erased. The roll-up mode distinguishes the captions as being real time.

The paint-on mode is currently used within the text service. The text is displayed as it is received from left to right beginning at the top of the screen and working down to row 15 after which the text scrolls up.

Captions and text may have any of several attributes in addition to block monochrome characters. Attributes include upper and lower case, six different colors, italics (or slanted text), underline, and flash. Attributes are determined by the editor at the time of encoding the captions and text.

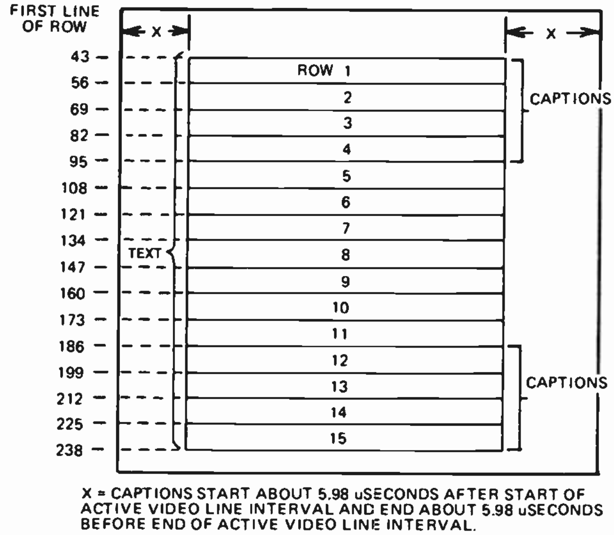


Figure 5.13-2. Caption row positions.

The caption display on the television screen consists of 15 rows with up to 32 characters per row as shown in Figure 5.13-2. A switch in the decoder permits the viewer to select captions or text. In the captioning mode, a maximum of four rows is used onto which each caption pops-on when prerecorded captions are received and rolls-up when the captions are live. The four rows can appear anywhere on the 15 displayable rows which occupy most of the screen area.

The first row starts at line 43 in each field. Each row occupies 13 lines of a field scan or 26 lines of the 525 lines. When the decoder is in the text mode the screen is nearly covered by the 15 rows of text, which scroll upward.

Each row of characters is displayed within a black surround box to enhance the readability against the normal video background. The box extends one character position to the left of the initial character in each row and one character to the right of each row. Partial or segmented rows also conform to this black box surround.

CAPTION DATA ENCODERS

Encoding is the process of inserting the caption data into the VBI of the television signal. The data contains the caption text in addition to positional instructions and display attributes (color and italics). The encoder is placed in the video path of the program to be captioned. There are two versions of the line 21 VBI encoder *smart encoder* and *simple encoder*.

Smart Encoder

The smart encoder is used to insert caption data into line 21 of fields 1 or 2 of the VBI. The smart encoder receives data through an RS-232 serial data port or via telephone line from an internal modem. Through the use of the smart encoder, locally produced caption and text services may be added to an already closed-captioned video program or if noncaption text is contained on line 21 (see block diagram of smart encoder in Figure 5.13-3). If captions are present on the incoming program, noncaption text may be interleaved into the gaps between the captions by the smart encoder.

Smart encoders are used in the process of creating real time captions, live display captions and captions from newsroom computers (prompters).

Simple Encoder

The simple encoder generates the line 21 data signal, to be inserted on a video signal, from caption data received at the RS-232 serial input. This encoder is used mainly for offline or prerecorded captions. The simple encoder cannot add captions to a video signal already containing line 21 data.

To create the line 21 caption signal, after the editor has produced the captions and they are stored in the computer, the videotape with program video and time code information is fed through the simple encoder (see block diagram of simple encoder in Figure 5.13-4). The time code information is used by the encoder to trigger the transfer of data (captions) from the captioning computer. The special software in the computer processes the data and matches each caption

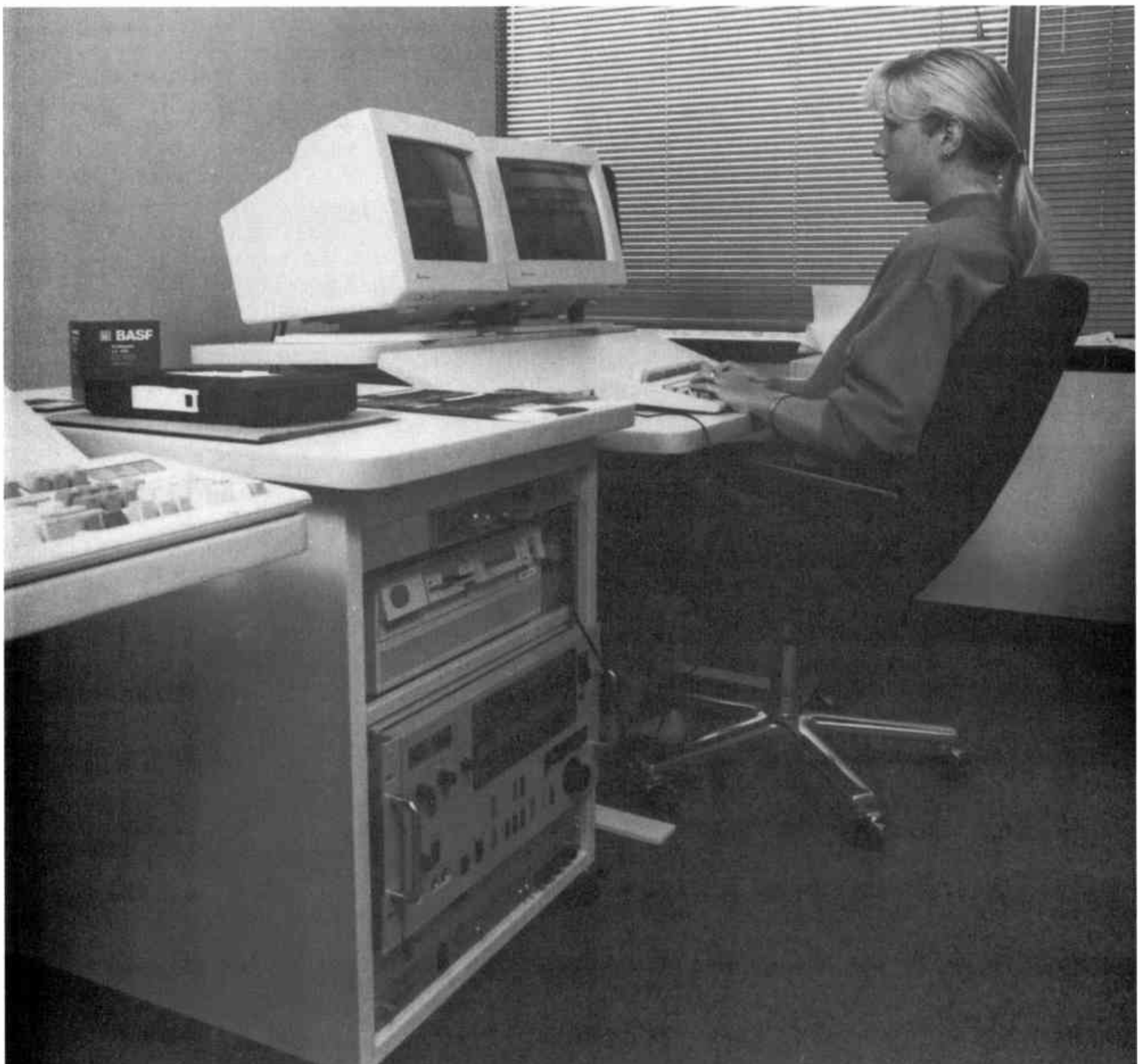
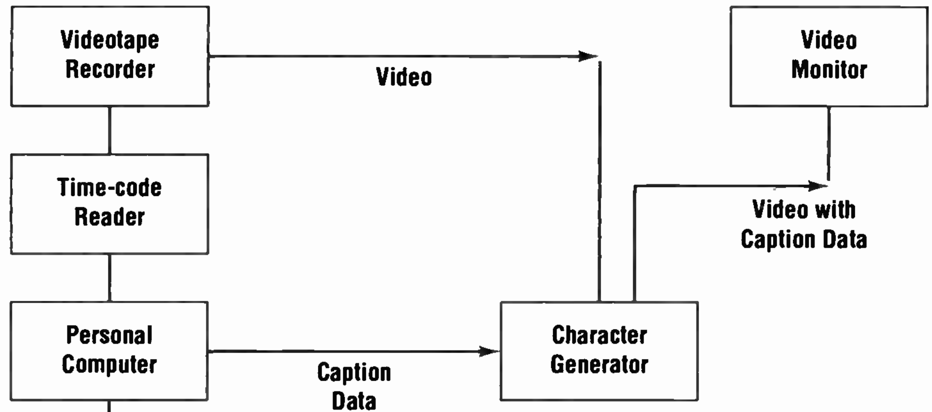


Figure 5.13-3. Captioning prerecorded materials is done by an editor working at a standard computer keyboard. (Courtesy of National Captioning Institute.)

Figure 5.13-4. Off-line captioning system. Caption creation.



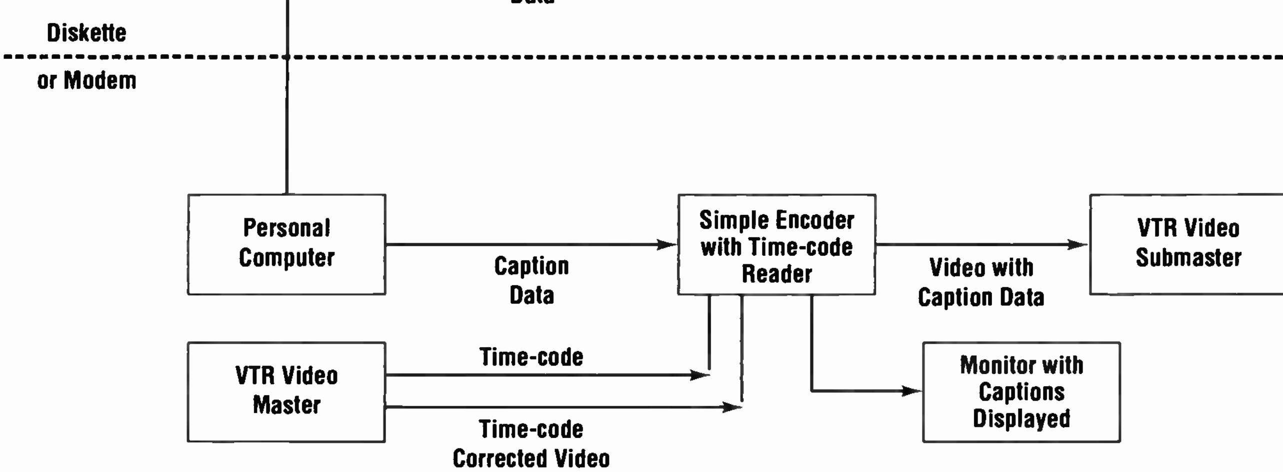


Figure 5.13-5. Off-line captioning system. Caption encoding.

to the appropriate time code. A microprocessor in the encoder processes the RS-232 data and encodes two characters of caption data onto each field 1 of line 21 of the program video. The output of the encoder may then be recorded on a second videotape recorder for playback at a later time.

Digital Encoders

In 1996, encoder manufacturers produced a new line of encoders to deal with analog and digital broadcasts.

METHODS OF CREATING CAPTION DATA

Several techniques are employed to create caption data based on whether the broadcast is live or prerecorded. There are also variations within the live and prerecorded captioning technologies.

Prerecorded Captioning

Prerecorded captioning (offline or nonlive captioning) involves the preparation of closed captions for programs that have been recorded prior to their telecast. The captions are also created in advance, and a closed captioned videotape version of the program is made. There are several steps in this process.

A captioning facility receives a time coded videotape copy of a program master. A caption editor reviews each scene, listens to the spoken words, and types the captions using a standard keyboard on a personal computer (PC) (see Figure 5.13-3). Several quality control checks are performed, including spelling, grammar, syntax, timing, and screen positioning of the captions. The captions are then stored with their corresponding time codes on a diskette. Figure 5.13-4 shows a block diagram of a captioning facility.

The recorded program and data from the diskette are then merged, using the time codes to trigger the captions, to produce a closed captioned submaster videotape recording, with the captions encoded on line 21 (see Figure 5.13-5). The encoded version of the program is returned to the producer and is used when the program is aired.

Captions which are produced in advance are displayed on a caption equipped receiver as pop-on captions, each one complete and timed to coincide with the spoken dialogue on the screen.

The equipment necessary for creating prerecorded or offline captions includes:

- PC with caption creation software
- Video monitor
- Audio monitor
- Character generator
- Time code reader
- Videotape player
- Simple caption encoder
- Videotape recorder

Using this equipment, the caption editor can view the program, listen to the audio, enter the captions along with the appropriate timing and placement information and review the program as it would appear through a decoder. It takes approximately 20–25 staff hours to caption one hour of television programming.

Live Captioning

Live captioning involves the addition of caption data to the television signal at the time of a live transmission or broadcast. Examples of live captioned programming include news programs, sporting events, news conferences and special bulletins or reports. There are three versions of live captioning.

Real Time Captions

For *real time captions*, a specially trained caption editor listens to the audio of a live television program and, using a stenotype machine similar to the ones used by court reporters, keys in words using a special shorthand code corresponding to syllables or phonetic codes (see Figure 5.13-6). Instead of keying in individual letters, groups of keys can be pressed down simultaneously. Each group of keys, a stroke, produces different phonetics. Using this kind of machine saves time, since a word such as *institute* may be keyed in using just three strokes (one for each syllable) rather than nine (one for each letter).

With a stenotype machine, the real time caption editor can key in up to 260 words per minute. The phonetic codes are translated into English words by the captioning computer which has been programmed with the phonetic codes and caption editor dictionary (see Figure 5.13-7). From the computer, the words and caption control codes are sent via a data circuit to

Figure 5.13-6. Real-time captioning is done by an operator using a stenotype keyboard similar to those used by court reporters (Courtesy of National Captioning Institute).

Figure 5.13-7. Live captioning system. Caption creation.

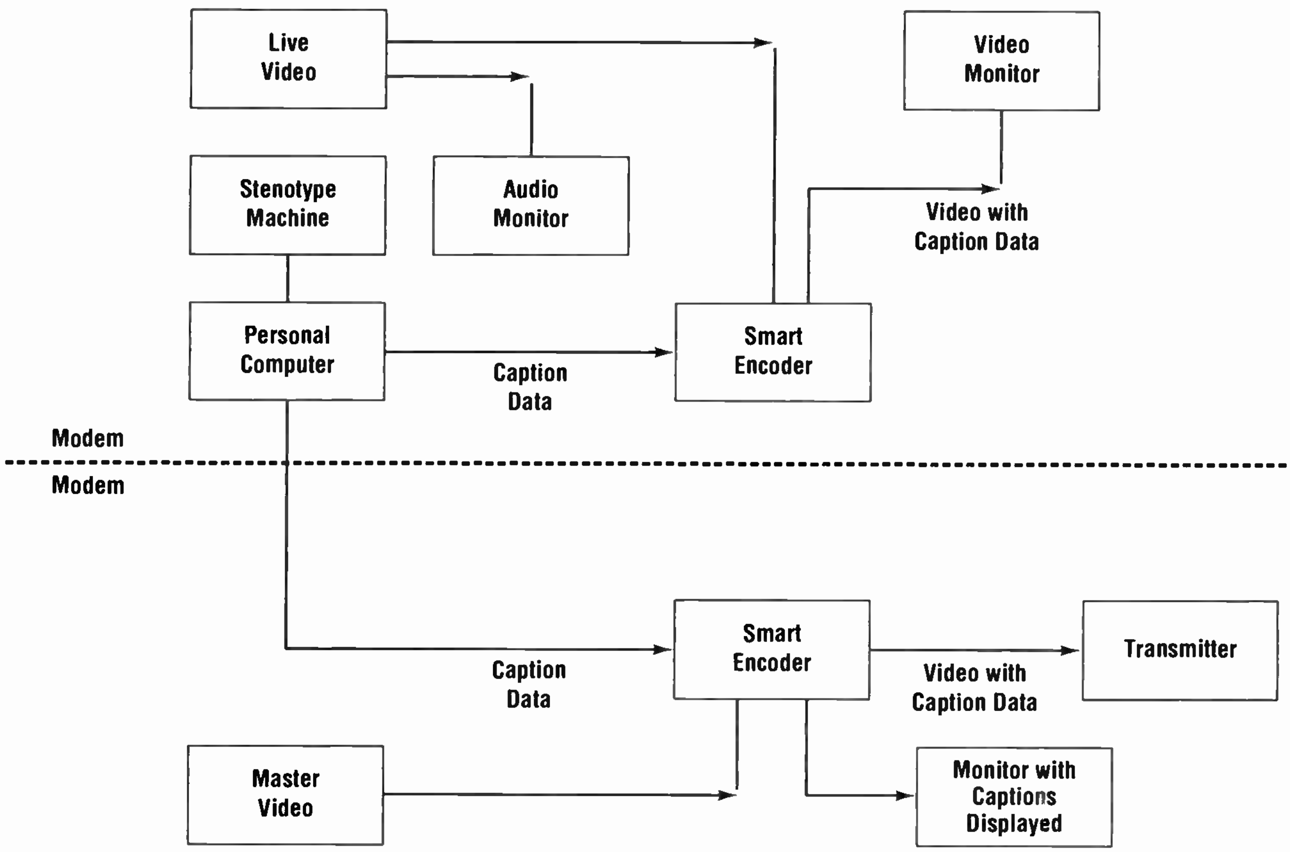


Figure 5.13-8. Live captioning system. Caption encoding.

a caption encoder where they are encoded into the television signal on line 21 as real time, or roll-up captions (see Figure 5.13-8). A television receiver with a decoder then displays these words as part of a continuous multi-line (1 to 4 lines) scroll (caption) at the lower portion of the television screen. It takes up to three seconds from the time a word is spoken to the time it appears on the screen as a caption.

Newscasters often speak faster than the captions can appear on the TV screen and may occur faster than the viewer can read. To keep up with what is being said, and to control the display rate, the real time caption editor must, at times, edit words or delete phrases. The real time caption editor must also realize that the computer dictionary may not contain a particular spoken word, and substitute a different word, one the computer will understand and at the same time convey the meaning to the caption viewer. Alternatively, if the computer does not have the word, the captions are transmitted as phonetically spelled out.

The equipment needed for creating and monitoring real time captions includes:

- Special stenotype machine
- Personal computer with translation software
- Video monitor
- Audio monitor
- Smart caption encoder
- Live video program signal on which line 21 will be encoded with the captioning signal

The real time captioning equipment need not be located where the live program is taking place. It is only necessary for the caption editor to hear and see the live program material (even over-the-air) and, using the stenotype machine, enter the captions into the captioning computer. The computer transmits the data through a modem to the caption encoder which must be placed in the program line of the live video signal to be captioned. Thus, any live program may be captioned as long as the caption editor has access to the live program audio signal and a caption encoder is in the live program circuit. Of course, it is desirable to be able to see the results of the captioning process and the same over-the-air signal can be used for the monitor.

For real time captioning of programs which may not be broadcast at the time of captioning or for which an over-the-air signal is not available, a microwave feed of the encoded signal may be used as the program monitor.

Because of the investment required for a captioning

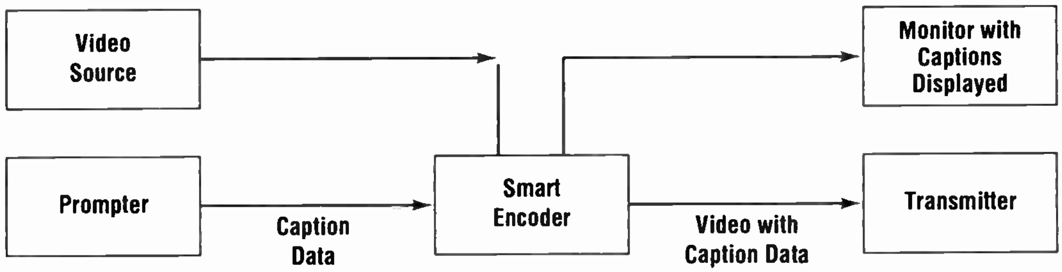


Figure 5.13-9. Captioning in an electronic newsroom.

facility and the high level of skill needed by the real time caption editor, stations and production facilities may find it more practical to employ the services of a captioning agency where captioning equipment and personnel are more readily available and the cost is based on an hourly basis.

Live Display Captions

Live display captioning is used when an accurate script is available in advance of a televised live event such as a speech or newscast. The scripted words are converted to captions by an editor and stored on a computer disk. When the live event is televised, the editor manually calls (triggers) each caption from the disk which is then processed by the computer and transmits the data to a caption encoder. The editor transmits a display command when each caption is to be displayed. With live display captioning, the words are timed to appear on the television receiver as they are being spoken. If last minute changes are made in the script it may be necessary to switch to real time captioning.

Newsroom Computer Captions

The third method of live captioning is the generation of captions through the use of newsroom computer equipment. Many television newsrooms convert news stories from their word processors into data for use on a prompting machine for the news reporter to read while on-the-air. At the same time the reporter reads the script from the prompter, the computer controlling the prompter passes the data through a serial interface to a caption encoder, through which the live program signal passes, that inserts the captions into line 21 of the video signal (see Figure 5.13-9). The result is the scroll or roll-up display on the viewer's screen.

There are some dangers in this process. It is not as automatic as might be assumed. Some newsrooms add cues to the prompting text designed for the on-air personnel, some of which might be inappropriate to appear as captions. Program breaks or changes of scene may require changes in the prompting operation that could interrupt the captioning process. Also, it is desirable for the operators and program producers to have an off-air monitor to insure the integrity of the captioning process. Finally, it may be necessary for an operator to signal the captioning encoder to add captions at the beginning of the local program and to revert to the pass through mode at the end.

The equipment needed for electronic newsroom captioning includes:

- Prompting system with data output
- Captioning software for the prompting system
- Smart caption encoder

Automated Live Encoding

When production schedules are tight, an alternate means of producing a captioned submaster is available. *Automatic live caption encoding* makes use of the same caption creation techniques used in prerecorded captioning, but a different method is used to trigger the data into line 21 of the television signal. Figure 5.13-10 shows a block diagram of the system. In this case, the captioned data are loaded into the PC and the internal clock within the PC is used to trigger the captions as opposed to using time-code from the program videotape. A manual trigger is used to start the transmission of data between the PC and the smart encoder.

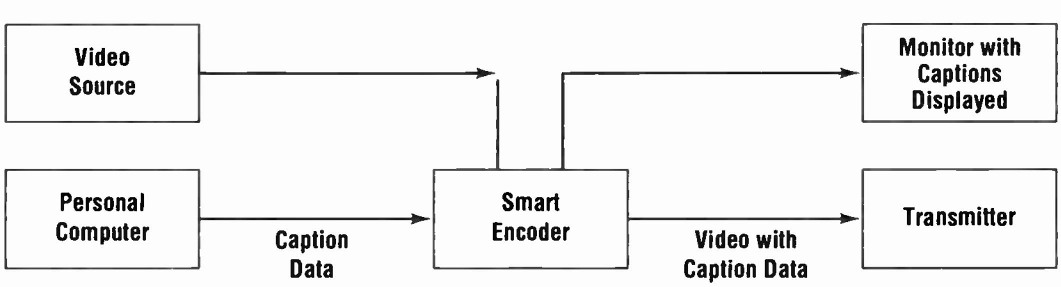


Figure 5.13-10. Block diagram of an automatic live encoding system.

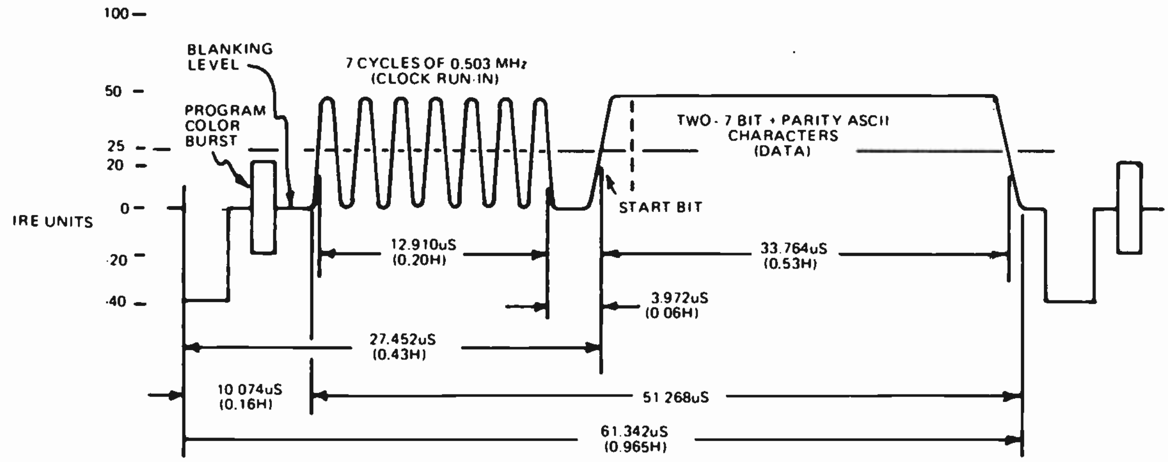


Figure 5.13-11. Line 21 field 1 data signal format.

The display of automatic live caption encoding on the television receiver is the pop-up display mode, the same as used for prerecorded captions.

Text Service

Information transmitted as part of the text service is usually independent of the program with which it is transmitted. Captions do not use all the available data capacity of the line 21 captioning system, and the additional capacity may be used to transmit nonprogram related information to the television viewer. Such text may consist of program notes, news, weather, sports, farm and financial reports and other information. Unlike the captions, the text service appears as full screen rolling text (without pictures) on the television screen.

The text service is accessed in the decoder by a switch that selects either captioning or text. There are two text channels for each field. Field 1 includes T1 and T2. Field 2 provides T3 and T4.

The data rate of the text service is quite low compared to the data rates of various teletext services that have been proposed for use in the United States and is in widespread use in Europe. The text may pause occasionally when the higher priority captions are being transmitted.

The equipment necessary for creation of a line 21 text service includes:

- a personal computer with appropriate text generation software
- an information source
- a smart encoder along with appropriate monitors

The text service can, under certain conditions, operate completely unattended if desired.

DATA TRANSMISSION FORMAT

Captions associated with a television program are transmitted as an encoded composite data signal during line 21 on field 1 of the standard NTSC video signal shown in Figure 5.13-11. The signal consists of a clock run-in signal, a start bit, and 16 bits of data corresponding to two bytes of 8 bits each (7 bit ASCII code plus one parity bit). Therefore, transmission of actual data is 16 bits every 1/30 of a second or 480 bps. The data stream also contains encoded information (control codes) which provides the instructions for formatting and changing the attributes of the characters to be displayed.

The clock run-in consists of a seven-cycle sinusoidal burst which is frequency and phase locked to the caption data clock. The frequency of 32 f_H (0.503496 MHz $= 32 \times 15{,}734.26$ Hz), which is twice that of the data clock, provides synchronization for the decoder clock. The clock run-in signal is followed by the equivalent of two data bits at logical zero level, then a logical one start bit. The last two cycles of the clock run-in, the two logical zero bits, and the logical one start bit constitute an 8 bit frame code signifying the start of data as shown in Figure 5.13-12.

The 7 bit ASCII transmitted data are coded in a nonreturn-to-zero (NRZ) format. An 8th bit is added to each character to provide odd parity for error detection.

The sequence of identification, control, and character code transmission is shown in Figure 5.13-13 and 5.13-14. Each caption transmission is preceded by a preamble code, which consists of a nonprinting character followed by a printing character to form a row address and display color code. Both characters of all control codes are transmitted within the same field of line 21 and twice in succession to ensure correct reception of control information. A transmitted caption may be interrupted by a mid-caption control code between two words in order to change display attributes such as color or italics. At the completion of a caption transmission, an end of caption control code is sent.

The first character of a control code is a nonprinting ASCII character (0000000 through 0011111). Codes 0000000 through 0010000 are not used. This is followed by a printing character (0100000 through

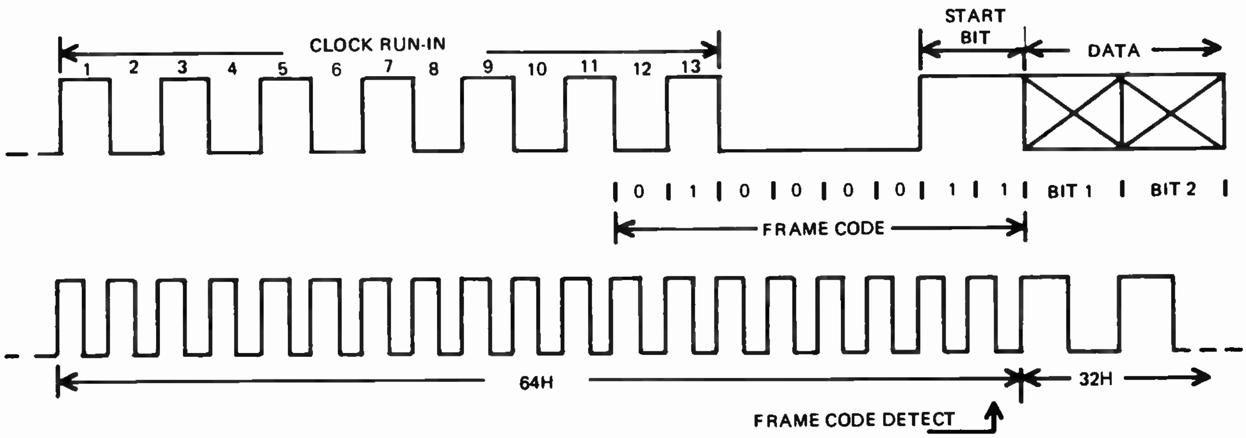


Figure 5.13-12. Line 21 data.

1111110). All characters that are received after a set of valid control codes are interpreted and loaded into the decoder memory as printing characters. Character codes with bad parity result in an all ones code being written into memory; this causes display of a white box (the delete symbol) in place of the desired character which, of course, was in error.

The data rate is 480 bps (8 bits per byte including parity). At an average of 5 bytes or letters per word, a maximum word rate of 12 words per second or 720 words per minute could be achieved. In practice, the word rate is somewhat less than this, about 600 per minute, due to the time required for transmission of control codes. Because most speech is much slower than this, there is adequate time for a second captioning channel and a text service.

LINE 21 TECHNICAL ADVISORY

Special precautions must be undertaken to ensure the continuous and correct passage of the line 21 signal throughout the facility. Unless a station is actively involved in producing captioned program materials, the line 21 signals may not normally be monitored at a television facility to the same extent as the video and audio signals. With the passage of the *Decoder Circuitry Act*, millions of caption receivers will be in use within a few years. Therefore, it is prudent for stations to install captioning monitoring facilities and to routinely check the path of the line 21 signal.

The captioning data signal is contained in the NTSC television signal VBI on line 21, field 1 (field 2 is not normally used for captioning but could be used for

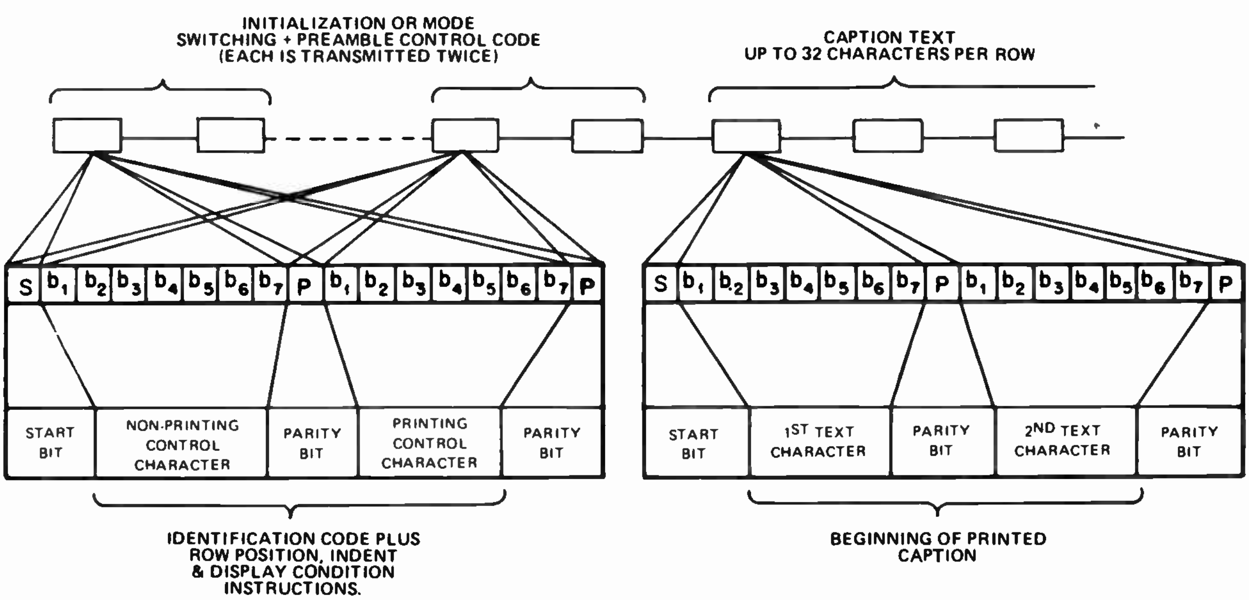


Figure 5.13-13. Caption row preamble format.

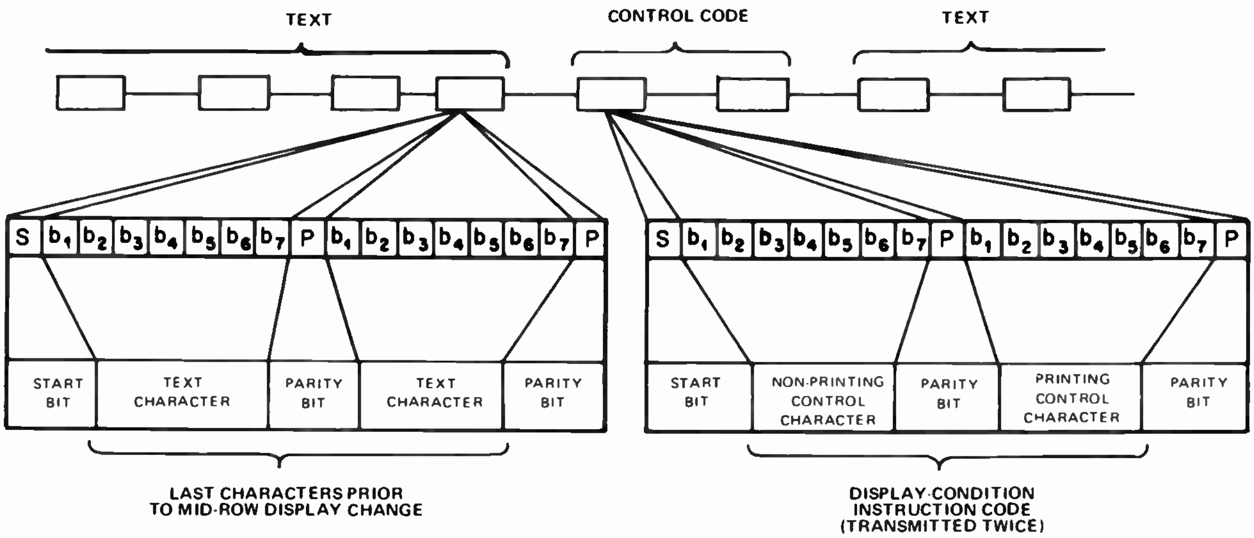


Figure 5.13-14. Mid-caption display. Condition change format.

text and signalling data). Unless the caption data signal is transmitted intact and on line 21, field 1, a viewer's decoding circuitry will not operate properly.

The captioning signal waveform is detailed in Section 73.682(a)(22) and Figure 17 of Section 73.699 of Part 73, Subpart E of the FCC Rules and Regulations and conforms to the standard television synchronizing waveform for color transmission given in Figure 5.13-6 of Section 73.699.

The captioning data signal is a part of the television program, compared to vertical interval test signals which are not. Regular observation and monitoring is essential in order to verify the presence and proper location of the data signal. Equipment throughout a video facility should be routinely checked and, if necessary, adjusted to pass the line 21, caption data signal.

Maintenance personnel as well as control room operators should be made aware of the importance of the line 21 signal and adjust all applicable equipment accordingly.

The following video processing equipment may adversely affect the line 21 signal:

- **Video processing amplifiers.** Certain types may delete the line 21 signal, move it to another line, exchange the fields or partially blank the signal.
- **Time base correctors (TBC).** Some may blank the line 21 signal or advance or delay line 21 by a field or a line.
- **Videotape machines.** Certain types contain a TBC which may blank or move line 21.
- **Video switchers or digital store devices.** May blank or move the line 21 signal.
- **Frame store or frame synchronizers.** May blank or distort the line 21 signal.
- **VITS generators or inserters.** Should be checked to determine if they are programmed to pass the line 21 signal correctly.
- **Digital effects generators.** May delete or distort the line 21 data signal.

A waveform monitor with line select capability is required to view the line 21 caption data to determine that it is on the correct line and field. For example, a Tektronix 1480 series; VM 700A, 1700 Series, 1780R, or Hewlett-Packard equivalent with line selector will easily locate and display the signal.

The line 21, field 1 data signal may be partially, but inadvertently, blanked. This can occur when the active video line is advanced or delayed relative to horizontal sync. The decoding system is more tolerant of this problem when the data signal is early and causes part of the clock run-in sinusoidal burst to be blanked. On the other hand, if the data signal is delayed relative to horizontal sync, loss of the last data bit (parity bit) may occur producing decoding errors which in turn will cause the characters to be displayed as white boxes. In an extreme case, too many errors will cause the decoder to disable all closed captioning functions.

Some video processing amplifiers have sensitive vertical phase blanking and horizontal blanking controls. These may also adversely affect the line 21 signal.

Video processing amplifiers are frequently used in the following configurations:

- Output of a satellite receiver
- Input or output of an intercity relay link (ICR)
- Input or output of a studio-to-transmitter link (STL)
- Input to a transmitter
- Input to a videotape machine
- Output of a studio or master control switcher
- Remote truck video processor
- Utility proc amp (equipment racks)
- Maintenance shop test equipment

All signal paths, backup transmitters and their VITS generators and processing amplifiers, and tape-delayed broadcast installations should also be routinely checked to insure passage of the line 21 signal.

It is important to note that if the overall signal level is allowed to drop substantially below the normal 50 IRE level, captions could become garbled.

Special Cautionary Notes

Locally Originated Captioning

Programs and commercials prior to, during, and after locally originated captioned programs may also be closed captioned. Therefore, the local smart encoder must be disabled, preventing pass-through, until the completion of a network telecast. The encoder must be enabled immediately following the locally captioned program to avoid stripping line 21 data from network programs which follow and pass through the caption data.

Newsroom Computers

Newsroom computer captioning typically uses the smart encoder in a newswire/real time mode. When this mode is invoked, incoming line 21 data are automatically turned off, preventing any previously recorded caption information from being passed through and displayed. The newswire/real time mode must be properly exited at the conclusion of local live captioning by issuing an *end of message* command to restore or enable the line 21 data.

Other systems may use the smart encoder pass-through mode rather than the newswire/real time mode. Pass-through offers more complete control of encoder and decoder functions. In this mode, incoming line 21 data can be specifically turned off and must be specifically turned back on (enabled) to allow line 21 caption information passing through the smart encoder to be displayed. The encoder must be specifically issued an enable command to turn on line 21.

A smart encoder reset command will always return the encoder to the default enabled condition but will delete any user stored articles, such as text channel information.

Network Origination/Local Origination

If a local station uses the master control routing switcher for signal distribution during a program originating from the network and then switches to local origination and a studio controller prior to the end of the network feed, the line 21 signal should be checked to insure its integrity.

Using Captioned Excerpts

Excerpts of programs, received or recorded off-air or from external satellite or network feeds, for use in other programming, may contain line 21 closed captions. In these cases, the captions should be stripped if the excerpted material audio will be edited or replaced because the captions may no longer be germane. A time base corrector or video processing amplifier can be used to remove or move the caption data signal from line 21.

Broadcast Decoder

In addition to consumer models, a professional broadcast type decoder is available for stations to use to monitor the line 21 data signal on the network, videotape machines, or the broadcast signal. The broadcast model accepts a standard video signal with captioning data on line 21 and produces a video output with the captioning data decoded and, using an internal character generator, displays the data as open captions on a standard video monitor.

Many stations have also installed Telecommunication Devices for the Deaf (TDD) in order to communicate with the hearing impaired television viewers.

Emergency Messages

Broadcasters which elect to transmit emergency messages are required to provide a readable version of the aural emergency message in order to inform hearing impaired viewers of impending weather or other emergencies (see Section 73.1250(h) of the FCC Rules). One means of doing this is to insert a moving line of words (crawl) with the message across the lower portion of the screen. If a program, over which the message is superimposed, is captioned the lower portion of the viewer's screen may be covered with the captions and the emergency message may become obscured or unreadable. Therefore, procedures should be established to stop or interrupt the captions or text transmission during the emergency message or to incorporate the message into the captions or text material.

Passage Through Other Distribution Systems

The line 21 captioning signal was designed to be robust and able to withstand a significant amount of degradation before errors appear in the text. The low data rate and placement of the signal at the very beginning of the active portion video of the signal contributes to this robustness. As a direct result, the captioning data signal will easily pass through:

- Master antenna (MATV) systems
- Community antenna (CATV) systems
- Satellite circuits and receivers
- Translators, multipoint distribution systems (MDS) and multichannel multipoint distribution systems (MMDS)
- Set-top converters (RF and baseband)
- Consumer video cassette recorders (VCR)

Most set-top receiving converters have outputs on channel 2, 3, or 4. A stand alone captioning decoder can be inserted between virtually any converter and any television receiver with little degradation to picture and sound quality.

5.14
FILM FOR TELEVISION

RICHARD W. BAUER
CONSULTANT, ROCHESTER, NY

INTRODUCTION

Until the development of the quadraplex videotape recorder in the 1950's, film was the only medium for recording television productions. Film was the primary distribution medium for the delivery of syndicated television programs to the local broadcaster for many years. It was used for news gathering and broadcast and local TV stations maintained large film libraries. The major network broadcast centers used *kinescope* recording systems and even some local stations had film processing facilities. The use of kinescope recorders to record TV images on film, broadcast telecines and videofilm news gathering are now of mostly historical interest. Direct broadcast of film is not common practice today but film still retains its important place in the origination of broadcast images.

To many broadcast engineers, the use of film in the broadcast studio is almost unheard of today. In the U.S., most film chains have been removed from the local facility. Film libraries, film projectors and film processing equipment are no longer in use. Although the presence of film and film related equipment is not obvious in the local TV studio, the images produced on film are still very important to the broadcaster and telecines are still in use at the network broadcast centers and in some of the larger facilities elsewhere.

The close relationship between film and television goes back than 50 years and it continues to grow as new films and new electronics continue to be developed. The motion picture industry supplied feature films for TV programs as well as camera people, craftsmen and technicians and this close association still exists today. In the early growth years of television broadcasting, the use of film for kinescope recordings allowed for the broadcast of programs on both coasts, for delayed broadcasting and for the distribution of programming worldwide. For example, programs originating in New York were kinescoped and the film was immediately flown to the other cities or other broadcast centers for broadcast. The electronic/photographic systems used produced poor quality images and, because of the very rushed nature of the work along with poor processing control and film handling techniques, a less than desirable end product was produced. The quality level of the various components of the system (CRT, film, film processing equipment, personnel) was not conducive to providing the image quality seen in the movie theaters or on TV from live broadcasts. This, however, was the only means of providing images for delayed broadcast and distribution. Now quadraplex recorders, trinoscopes and broadcast telecines are of more historical interest rather than day to day operations. The local broadcaster no longer needs to maintain a film library, broadcast telecine or other film related peripherals.

The use of film for kinescope recordings caused many broadcast people to denigrate film and its fuzzy, grainy image. These feelings were well founded as far as the finished *kine* was concerned but the criticisms were somewhat misdirected at only the film. Film manufacturers made film for the motion picture industry, not for the fledgling TV industry. However, as TV grew the film and television engineers worked closely to develop systems that would utilize the best components from both industries to provide the high broadcast image quality enjoyed today. Films were specifically designed for TV and reproducing equipment was specifically designed for film.

When TV was just converting to color broadcasting, color motion picture film was ready and produced the color programs. However, news coverage was still restricted to black and white images from the video news cameras. The Eastman Kodak Company developed a color reversal, high speed news film with integral magnetic track for single system sound and image recording. This system was the basis of much expanded, rapid news coverage on the local and national level and flourished until the development of portable color recording video cameras and tape recorders.

Since those early years, the films have been improved for speed, grain, sharpness and contrast, all of which make the reproduction of film images on television better. Also of high importance are the developments in flying-spot and charge coupled device (CCD) telecines, automatic color correction and the digital revolution now in progress. With the development of high definition television (HDTV) film will continue to provide the quality images as it meets or exceeds the high line resolution and it provides a standardized format that can be used with any system chosen.

In the U.S. and internationally, film remains the preferred medium for the origination of television programs and commercials. Motion picture film in 35mm width, four-perforation pull-down, and 24 fps projection is the worldwide standard. It can be projected or transferred via telecine to any videotape format or line scan. Film has maintained its role as the primary production medium for prime time television programming. The percentage of prime time programs originated on film has remained fairly constant, above 80%, for nearly 40 years. With the many improvements in telecines, electronic editing, auto-color correction and digital tape recording, many of the programs are directly transferred from the camera negative film and post-produced on videotape. Many national commercials are also shot on film then post-produced in the same way.

Film and television are an inseparable combination that continues to grow and prosper. As film characteristics are improved for TV, the electronics are challenged to reproduce the improved quality, thus leading to the development of the next generation of high-quality systems. Then the cycle starts all over again. It is this relationship that has greatly contributed to the success and popularity of modern day television.

FILM CHARACTERISTICS

Color motion picture film has three photosensitive layers, sensitized to red, green and blue light. Exposure and processing produce cyan, magenta and yellow dye images in these layers which, when projected on a large screen or scanned in a telecine, produce the color pictures. Although the 16mm width was the most common format for distribution of television programs for many years, 35mm is currently the most used format but *super 35* and *super 16* are also utilized.

Film (35mm) is the main format for large screen projection in theaters *and* is also used for the production of prime-time television programs. This format is used extensively in the post-production houses and for the production of television commercials. Film continues to be improved for speed, grain and sharpness by the manufacturers. Eastman Kodak Company estimates another 10× improvement is still possible with silver halide technology. The quality of film images, versatility of production techniques and worldwide standardization keep 35mm film an important and valuable tool for the production of images for television broadcast. New versions of both flying-spot type and CCD film scanners continue to be developed for direct broadcast or transfer of film to video. Television film scanners have interchangeable optical blocks enabling either 16mm or 35mm format to be reproduced in telecine equipment.

Professional motion pictures, most prime-time television programs and television commercials are originated on color negative film. From these originals, prints can be made for broadcast and distribution on either 35mm or 16mm formats. Duplicate negatives can be made and large numbers of prints prepared for theatrical release. Telecines are capable of scanning either color positives or color negatives. A program originated on color negative can be transferred directly to tape, edited electronically and then broadcast. The same negative can be edited and either broadcast from film or a transferred tape. There are many choices available when the program or commercial is originated on motion picture film.

To display motion pictures on large screens in the theater, the projector advances the film one frame at a time and a rotating shutter cuts off the light while the film is being moved. A *claw* or *Geneva mechanism* advances the film by engaging perforations along the edges. Film 35mm wide has perforations on both edges, but the 16mm material used to make prints has perforations on one edge only; the space on the other edge is taken up by the sound track.

In camera tube telecines, the film advances intermittently one frame at a time by means of the perforations just as in theater projectors. But in film scanners, the film is moved continuously and the perforations are used only to drive a sprocket that generates framing pulses.

The film sound track may be either optical or magnetic, although magnetic striped film is rarely used today. Telecine projectors and film scanners usually have interchangeable optical/magnetic sound playback heads. In the playback of an optical sound track, the lens imaging the exciter lamp filament at the film plane must be sharply focused on the emulsion side. Usually an adjustment is provided for this purpose, as the emulsion side in 16mm prints may be either towards the lens (preferred position) or towards the light source.

The standard motion picture frame rate worldwide is 24 fps for both 35mm and 16mm formats, although 25 fps is often used for 50 Hz TV systems like PAL. The 35mm format has 16 frames per foot so at 24 fps, the rate of film movement in a projector or scanner is 90 feet per minute (fpm). With 40 frames per foot in 16mm format, the film passes through a projector or scanner at 36 fpm.

The super formats, super 16 and super 35, provide larger picture areas by using the sound track area for picture, with double system sound on another media.

When motion picture prints made for theatrical projections were a primary source of program materials for broadcasting, significant quality losses occurred. This was due to the inability of the television system to transmit the range of gray scale value needed to create acceptable pictures on motion picture screens. Committees of the Society of Motion Picture and Television Engineers (SMPTE) addressed this problem and published a report recommending the light density range in films for television be limited by adjustments in staging and lighting, rather than in the making of color prints. In this way, prints could be produced with a range of density values that the television system, at that time, could reproduce.

For creative and practical reasons, it was not always possible to adhere to the recommended lighting ratio of 2:1; nor to the recommended 60% reflectivity of the lightest scene element. Market requirements that made necessary the production of motion pictures for both purposes (theater projection and television display) have since changed, but at the time were extremely important.

The problem was addressed by the Eastman Kodak Company with the development of a low contrast print film. This film was designed with reduced upper scale contrast compared with the print film for projection on theater screens. The transfer characteristic of the low contrast print film was, of necessity, a compromise among several considerations. The contrast had to be as low as possible for good telecine performance, yet high enough to provide acceptable screen images in direct projection in review rooms. Also, the color saturation had to be high enough to maintain a chroma gain at a fairly low level. When a film with a contrast range of 160:1 is being squeezed into the television system, the shadow areas in the pictures are compressed and much of the shadow detail is lost. Some improvement was achieved electronically through the use of black stretch circuitry in the telecine. Further improvement was made by using low contrast print film for the television prints.

Some modifications were made in various films and in lighting techniques but these compromises were not the whole answer. New films have been completely designed specifically for television program production and much better image quality results. Today movies are produced especially for television, so the films, lighting, sets and actions are all optimized for video transmission. With the new design telecines and films, compromises are no longer required. Optimum image reproduction can be obtained for large screen theatrical release, television broadcast and videotape release.

There have been and continue to be significant cooperative research and development programs among the telecine manufacturers and film manufacturers. Continued improvements in all elements of the film system, from the negative through the intermediates to the final print, contributed to major quality improvements in the broadcast of film on television. The design of circuitry to allow the use of negative working films for either transfer or broadcast also led to major quality improvements as well as economic gains for the use of film in television.

Film for Television

Many programs are still originating on film though most broadcasters today do not see the film itself, but rather a videotape transfer. Nearly 80% of the prime-time programming on U.S. television originates on film, whether the program is a dramatic or comedy series, made-for-television movie, or theatrical motion picture. A large percentage of commercials also originate on film but are transferred to tape for distribution and broadcast.

There are many reasons why film remains the medium of choice for origination. One of the main reasons is an undefined phenomenon called *film look*. This characteristic has defied quantification by performance parameters but continues to be a major consideration. Another advantage of film origination is the standard format worldwide. Programs originated on film today can be readily syndicated for distribution in today's 525 or 625 line standards or any of tomorrow's HDTV systems.

With today's family of color negative films, the cinematographer has the flexibility to choose high speed emulsions or slower, finer grain emulsions simply by changing the camera magazine. A comparison of film characteristics and video camera characteristics indicates that film still has the advantage for sensitivity and speed, tone scale reproduction and resolution.

Recently, the Eastman Kodak Company designed a film specifically for television production. Its speed rating is EI 640 for tungsten balanced light sources. This film has some unique features that allow the capture of more information in the origination negative that can be subsequently scanned and manipulated to give the final image qualities desired. The film has three features that take advantage of a telecine's controls: lift, gamma and gain. First, the film has a low contrast capture which by definition is a very long latitude film (~ 10–12 stops). The low contrast capture also allows outstanding highlight handling capability. Second, the film was designed to match the spectral response of modern telecines, which enables the telecine to see the red, green and blue color records neutrally. Third, the *toe*, or region of the film which captures shadow detail was optimized to be controlled by the gamma control. This allows better control of the amount of shadow detail and fleshtone contrast. Much better independence of control between the lift, gamma and gain is also achieved with this specially designed television production film. This film offers all of the benefits of conventional theatrical films, but with a design cooptimized for making the production of television programming faster and more efficient.

Motion picture film offers significant advantages as a medium for the production of TV programming, while videotape offers advantages for the handling and broadcasting. Films transferred to videotape can be broadcast and handled many times without losses due to dirt and scratches that may occur with repeated handling of film.

The Broadcast Telecine

Traditionally, telecines were used for the transfer and direct broadcast of syndicated television programs and feature films. Only print films were transferred, and generally without supervision with no scene-to-scene color correction other than that provided by automatic gain or automatic black level circuitry. Today all types of films—negative, intermediate or print in black and white or color—are routinely transferred to video.

The original broadcast telecines evolved from television camera technology combined with motion picture projectors. During the 1970's, flying spot and CCD telecines were developed for broadcast use. Now in the United States, telecines are rarely used for on-air broadcast, but some broadcast facilities maintain a telecine to transfer print film from their libraries to videotape for local broadcast.

The Transfer Telecine

The transfer telecine is used to transfer feature films to videotape masters for subsequent duplication to videocassettes and videodisc formats as well as for broadcast. Flying spot and CCD telecines are used by film-to-tape transfer houses in this market. A skilled colorist provides scene-to-scene color correction, using programmable color correction controls. The transfer telecine is designed to accommodate negative film, print film, low contrast print film, duplicate negative or interpositive film. Some television programs are edited on film and transferred to videotape in a similar manner with scene-to-scene color correction.

The Production Telecine

The production telecine is used in the production of television programs and commercials. Post production facilities use flying spot and CCD production telecines. Selected camera shots (*circle takes*) are transferred from the camera negative film to videotape with a colorist providing scene-to-scene color correction.

Commercials may be edited directly (online), but television programs are typically edited offline. Dubs are made to videodisc or videocassette formats for offline editing. The edit decision list (EDL) generated in the offline edit session is then used to assemble a finished videotape master in an online editing session.

For internationally syndicated television programs, a PAL master is made. Video standards conversion from the NTSC master has proven inadequate, due to compromised spatial resolution and motion judder. The preferred approach is to cut the negative film to a cut list generated from the EDL that was produced in the offline edit session. The cut negative film is then transferred directly to a PAL master with scene-to-scene color correction. The recently introduced machine-readable edge numbers (Key-kode™) are expected to facilitate the process of negative cutting.

TELECINE DESIGN

The basic function of a telecine is to convert an optical image on motion picture film to a television signal. This conversion involves an optoelectronic transducer and a scanning operation. The resulting video signal must be color matrixed, gamma corrected and frame rate converted.

There are two basic telecine designs that are commercially available: cathode ray tube flying spot scanner (FSS) and CCD line array. An experimental laser telecine for HDTV has also been demonstrated by NAC.

A third design which was actually the original system used was the *photoconductive telecine*. This design involves the combination of a synchronized motion picture projector with a television camera. As the design evolved, specialized video signal processing circuitry was developed to improve the image quality of the transferred or broadcast images. The photoconductive (camera tube) telecine is no longer manufactured and probably only a few still exist. It was the first system and it was manufactured by a large number of companies worldwide, but with the development of the flying-spot and the CCD line array systems the photoconductive system could not compete for quality of the reproduced images nor with the ease of operation and maintenance.

Figure 5.14-1 illustrates the major components of a generic telecine design. These components include:

- film transport
- illumination system
- projection lens
- scanning system
- video signal processing

Implementation of these basic components depends on the telecine technology and design.

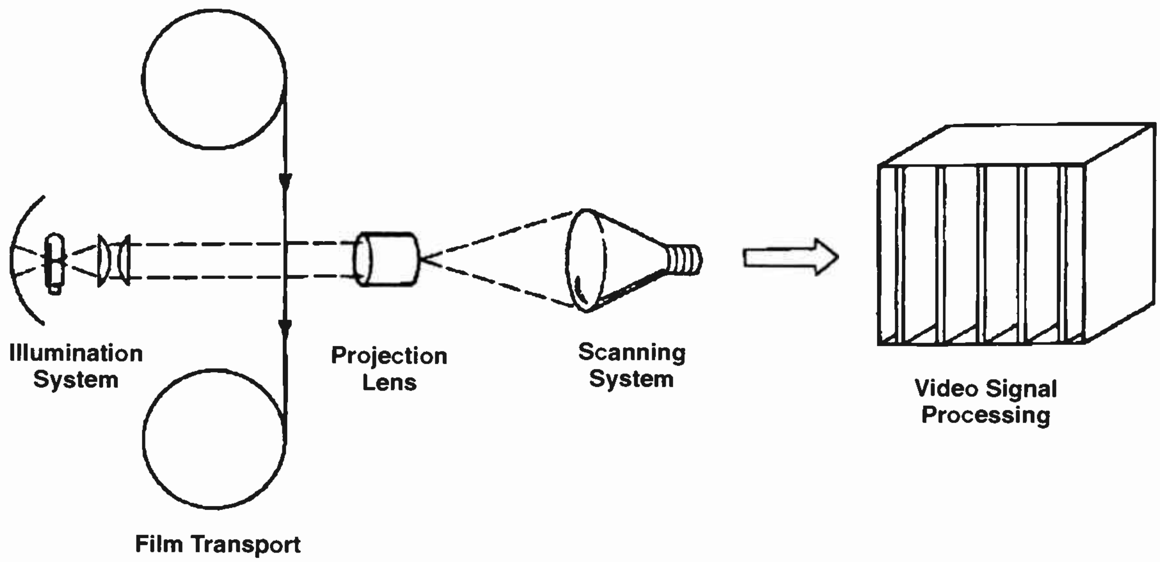


Figure 5.14-1. Generic telecine components.

CRT Flying Spot Scanner

The CRT FSS produces the video signal by scanning the film images with a very small spot of light and collecting the light transmitted through the film with a *photomultiplier tube* (PMT). A high intensity CRT is scanned by an unblanked electron beam.

The CRT FSS was developed by Rank Cintel and originally designed for 25 fps film transfer to 25 fps European television standards (PAL and SECAM). Early attempts at 30 fps NTSC designs involved a complicated *jump-scan* approach, controlling the scan to implement both interlace and 3:2 frame rate conversion. The development of the *Digiscan* frame store, which permitted the film frame to be progressively scanned and then interlaced, and frame rate converted by controlling the output (read) rate, made the FSS design practical for NTSC.

One of the fundamental innovations of the FSS design was the development of a continuous motion, capstan driven film transport. The velocity of the film is monitored by a shaft encoder on a free-running timing sprocket that tracks the film perforations, and the timing pulses are used to control the capstan velocity via a servo loop. This transport has proven to be gentle enough to handle negative and intermediate film stocks in addition to print film.

One of the fundamental challenges in the CRT FSS design is the selection of a high intensity, short persistence phosphor with a broad spectral energy distribution for scanning color film dyes. The typical phosphor has little blue output, limiting the signal-to-noise ratio of the blue record of transfers from negative film, as negative film contains yellow masking couplers (0.90D).

The basic block diagram for a CRT FSS is shown in Figure 5.14-2. The scanning spot is divided into three color channels by a dichroic beam splitter where each signal is picked up by a PMT. The signal from each PMT is buffered by the head amplifier and applied to the afterglow correction circuitry. Afterglow correction is a high pass filtering operation that compensates for the persistence (afterglow) of the phosphor.

The next step is shading correction, which compensates for uneven illumination, optical losses and uneven tube sensitivity. Here, shading correction compensates for nonuniformity of the CRT and lens system.

The video signal processing is implemented in four steps: (1) log amp, (2) color masking, (3) variable gamma and (4) expo amp. Color masking and gamma correction are implemented on logarithmic signals and the resulting signal is exponentiated for display. The color masking operation is implemented by a resistor bridge matrix, with selectable options for different film stocks. Electronic color masking compensates for the unwanted absorption, or crosstalk between the film dye records and the spectral response of the telecine.

The variable gamma function is implemented as a gain adjustment on the log signal. Black level (or lift) correction is applied by controlling the reference pulse to the log amp, and white level (gain) correction is applied at the expo amp.

A digital frame store is used to provide both interlace conversion and frame rate conversion (24 fps to 30 fps). This is accomplished by independent control of the input (write) clocks and the output (read) clocks. Aperture correction is also implemented digitally.

Recent advances in CRT flying spot scanners include an all digital video signal processing channel. Also of note is the development of a pin registered gate for image compositing and the development of an electronic pin registration (EPR) system for real time steadiness correction. In modern telecine equipment practically all of the analog signal processing has been replaced with digital signal processing. This has improved the reliability and the creative flexibility.

The primary advantage of the CRT FSS is the scan flexibility—zoom, pan and anamorphic expansion are

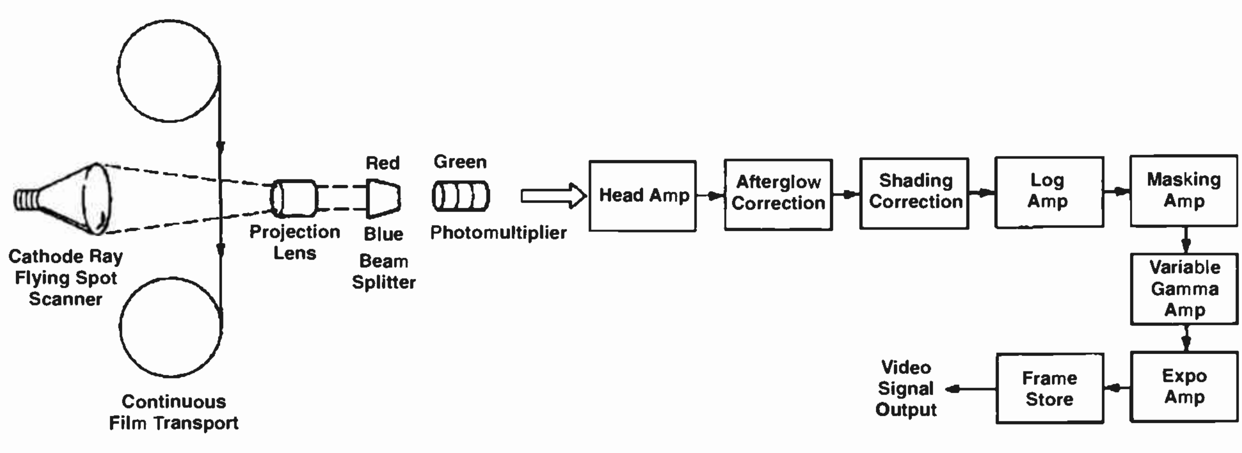


Figure 5.14-2. Block diagram of cathode ray tube flying spot scanner.

easily implemented by changing the scanning raster. The continuous motion transport handles film gently, making the transfer of camera negative film viable. Also, the lag and color misregistration artifacts of photo conductive telecines are eliminated.

The limitation of the CRT design is primarily the short life (2,000 to 5,000 hours) of the tube before it must be replaced due to phosphor burn or spot size deterioration. While not a real problem with NTSC or PAL transfer, prototype CRT HDTV telecines exhibit limited sharpness and signal-to-noise performance (particularly when scanning negative film).

CCD Line Array

The CCD line array telecine was first introduced in the early 1980's. As its name implies, the heart of the system is a CCD line array imager which converts the optical image to a video signal by transferring the charge accumulated in each photosite of the line array through a charge coupled output register. CCD telecines are manufactured by Philips BTS and are used in production and broadcast applications. The CCD telecine design also utilizes the digital frame store and continuous motion transport first implemented in the CRT FSS.

The illumination system is a high energy, blue rich tungsten halogen lamp, which is important in achieving high signal-to-noise performance when transferring orange masked color negative films.

Sensor clocks are generated to control the integration time of each line and the pixel readout rate. The commonly used CCD element photosensor has dual channel readout with alternate samples interleaved.

Pattern correction removes any stripe patterns resulting from photosite sensitivity variations and output shift register mismatches.

The video signal processing in a CCD telecine is much the same as that of the CRT FSS telecine described earlier. A basic block diagram is illustrated in Figure 5.14-3.

Film-to-Video Transfer

Most programs or features which originate on motion picture film will be transferred to a video format. In addition to cinema theater release, a feature film may be released for distribution to airlines, television broadcasters and home video retailers. Programs and commercials which originate on film, such as documentaries, prime-time comedies and dramas, are transferred to video as part of the post production process.

Film continues as an origination source for television for a variety of reasons. Feature films are a tremendous and popular resource of program material for broadcasters. Film creates images with the unique film look due to the inherent way in which motion picture film captures and records image records. Film is a worldwide standard and archival medium which continues as a significant image source for the television program production.

Telecine equipment makes the job of transferring film to video fast and efficient. Telecine equipment also provides the opportunity for creativity as the operator has the option to interact with the telecine controls to change image characteristics such as color, contrast, sharpness, zoom, and crop.

For the conversion of feature films, a telecine converts the film signal to the proper video signal and aspect ratio. Often feature film releases are shot in wide aspect ratios such as 2.1. The standard definition 525 and 625 video standards requires an image with an aspect ratio of 1.33. A telecine will produce an image with an aspect ratio of 1.33 (or 1.78 for wide screen or HDTV) from the film image record. This can be accomplished by *pan and scan* or *letterboxing*. During pan and scan a full resolution, smaller aspect ratio image is extracted from the larger aspect ratio film image. During letterboxing, a lower resolution, wide aspect ratio image is placed in the 1.33 aspect ratio video image. Black bars are added to the top and bottom of the image.

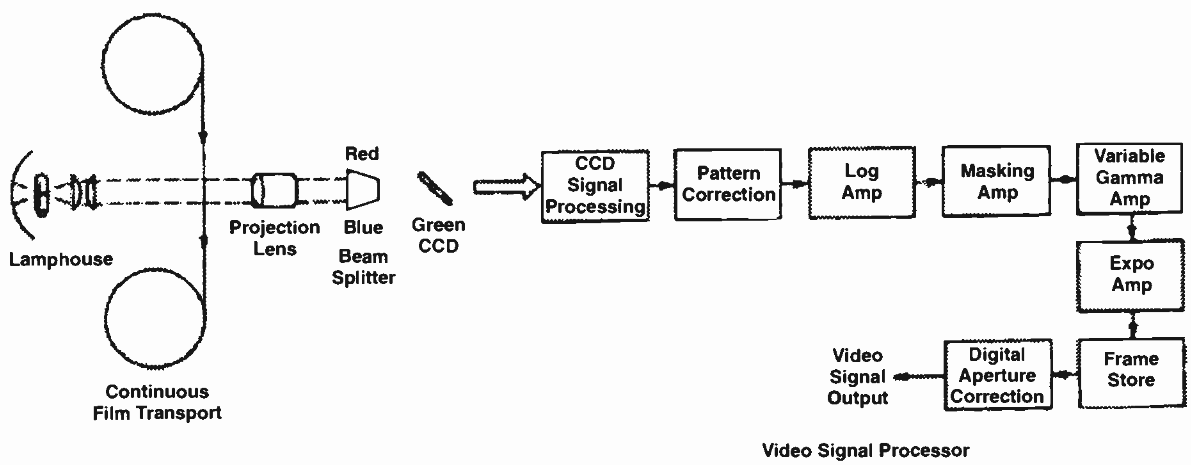


Figure 5.14-3. Block diagram of CCD line array telecine.

Telecine manufacturers make a variety of optical and transport gates to accommodate the film-to-video transfer of the various image formats which can be captured on 8, 16, 35 and 70mm film formats. Telecines handle the various film types such as original camera negative, reversal, black and white, intermediate and print film types by automatic alterations to the optical configuration and electronic signal processing.

For a given television program, the actual film material which will be transferred to video will depend on actual client preferences, requirements and material availability. Most motion picture film productions are captured on a camera negative material as this film material gives the cinematographer the greatest latitude in terms of capturing details and highlights and is the master version for making additional film duplicates. Some applications may require origination on reversal film, a film that produces an optical positive directly, without subsequent printing.

The film-to-video transfer may also be made from one of the subsequent film versions which originate from the camera negative, either intermediate film or print. Subsequent film generations exhibit different image quality features as the photographic process alters the characteristics of the image records. In order to optimize quality, the trend in film-to-video transfer has been to transfer from the original negative or from a film generation as close as possible to the original negative film.

Certain telecine machines also react differently to various film types. Some telecine equipment cannot handle the wide density range contained on print film. Eastman Kodak has produced a special low contrast print film for this application to reduce the adverse effects of print film transfer on these limited dynamic range telecines.

Telecine machines are very flexible in the types of video output obtainable. Practically all telecine machines can output both 525 and 625 video standards. Telecines output the variety of standard definition formats and are rapidly advancing to include high definition video images and digital computer data formats such as DXP and TIFF.

Programmable Color Correction

Telecines are controlled by programmable color correction devices. With these devices, the controls on a telecine can be programmed on scene by scene or frame by frame basis. With a programmable color correction unit, the telecine can be programmed to compensate for film differences contained within a roll or create special effects such as color enhancement, day for night, rotation, zoom, pan, noise reduction and sharpness enhancement.

Programmable color correctors can perform primary and secondary color corrections. Primary control refers to lift, gamma and gain controls which control the blacks, midscale and whites of an image respectively. Primaries also allow for red, green and blue interactions in the black, midscale and white portions of the image. Primaries are used to make overall color adjustments to the film transfer for the purpose of matching various elements or creating a color effect.

Secondary color correction refers to additional, precise manipulations. For example, a specific shade of red in a particular region of the image may be altered. Primary controls on a color corrector can be applied to both the signal processing within the telecine and the video processor of the color correction unit. Secondary controls are applied to the video processor of the color correction unit and usually cannot affect the internal workings of the telecine.

Programmable color correctors provide a multitude of other options for image manipulation such as processing of key and luminance channels. These devices control telecines and a variety of video recording tape machines and disk arrays.

MONITOR SETUP

The color picture monitor is the subjective reference for all film-to-video transfers. Although a video waveform monitor and vectorscope provide some objective signal levels, the color picture monitor is the basis for the artistic judgments of the colorist and the acceptance of the finished product by his client. Program interchange between facilities and consistent quality can only be assured by high standards for color picture monitor performance, alignment and viewing conditions.

The SMPTE Working Group on Professional/Studio Monitors (T14.28) has worked to develop recommended practices and standards for color picture monitors. These include:

- *Recommended Practice for Critical Viewing Conditions of Color Television Pictures (RP 166-1995)*
- *Recommended Practice for Alignment of Professional Television/Studio Color Picture Monitors (RP 167-1995)*

Performance

Performance criteria for color picture monitors identified in the SMPTE standard include the following specifications:

- White point = 35 ft L
- Color temperature = D65
- Colorimetry = SMPTE "C" phosphors
- Contrast ratio $\geq$ 50:1
- Gamma = 2.2 + 0.4/−0.0
- Luminance nonuniformity $\leq$ 25%
- Color purity $\leq 6\Delta$ E* CIELUV units
- Resolution $\geq$ 500 TV lines per picture height (without aperture correction)

Aperture correction should be switchable. Underscan capability is a must in order to see the full active video raster. Long and short-term stability requirements are also specified in the SMPTE standard.

Alignment

A color picture monitor can be aligned with a test signal generator and a color analyzer (*tristimulus device*). The required test signals will be indicated in the following discussion of alignment procedures.

- *Scan size* is adjusted in underscan mode so that all four corners of the raster are visible. Overscan mode should involve no more than 5%.
- *Geometry, linearity* and *aspect ratio* are adjusted using the crosshatch test signal, turning on only the green beam, and adjusting pin-cushion and scan-linearity controls for visual alignment with a linearity overlay (ball chart) over the CRT face.
- *Convergence* is adjusted with the crosshatch signal with all three beams on. The manufacturer's recommended adjustment sequence should be followed.
- If *aperture correction* is used, the amount should be set to make the apparent brightness of the 2T $\sin^2$-pulse match that of the bar in the *pulse-and-bar* signal. Alternatively, the multiburst signal can be used to match the apparent contrast of the 4.2 MHz burst to that of the 3 MHz burst.
- SMPTE color bars are used to set *chrominance amplitude* and *phase* for the decoder, viewing the blue channel only. The left blue bar is the reference bar and its brightness is not affected by the chroma or phase controls. The phase control affects the brightness of the inner two bars. The recommended adjustment sequence is to first adjust the chroma control and then the phase control so that all bars are of equal brightness (matching).
- A 100 IRE window signal is used to set *color temperature* and *reference white*. First a visual comparator is used to adjust the RGB gain controls for a visual match to D65 color temperature. A photometer is used to set the contrast control for a reference white level of 35 ± 3 ft.L.
- *Gray scale tracking* is set using an unmodulated stairstep signal and adjusting the RGB screen controls to produce visual neutrals in the darker steps. As the screen controls and gain controls interact, this requires an iterative approach.
- The last step is to use the Picture Line-Up Generator (PLUGE) signal to set the *brightness* control so the darker patch just merges with the reference black level, but the brighter patch is clearly distinguishable at the normal viewing distance. The PLUGE signal of the SMPTE color bars can be used.

Viewing Conditions

The recommended viewing distance for the telecine colorist is four to six picture heights. The monitor surround should be a neutral gray illuminated by D65 lighting in order to permit unbiased color judgments. A graded illumination is recommended to minimize fatigue, with a peak luminance of less than 3.5 ft.L (10% of reference white). A surround area of at least eight times the area of the color monitor is required. This can be obtained either by lighting a rear wall behind the monitor or by placing the monitor in a rear illuminated panel.

All room lighting should be filtered to D65 color temperature and directed so that there are no spurious reflections from the monitor face.

Telecine Exposure Calibration (TEC) Film

With the increased use of computer based editing systems, video dailies are replacing film dailies to a great extent. Processed negative is sent directly to the telecine transfer facility. Depending on post production needs, video dailies may not provide accurate feedback for directors of photography (DP). Without film dailies, there is no laboratory report with its list of printing lights to tell the DP whether the exposures are satisfactory before shooting the next setup. A telecine exposure calibration (TEC) film enables the colorist to calibrate the telecine and provide the DP with transfer points (TP), similar to the printing lights traditionally provided by laboratories.

The problem was how to calibrate an extremely complex device, such that it would give the same meaningful exposure data, no matter which telecine was used, which facility house was using it or where it was located. The solution, a TEC film is a strip of EASTMAN EXR color negative film, exposed and processed to precisely known densities. This film calibrates the telecine, turning it into a rather expensive color analyzer and densitometer, at least for exposure evaluation. The cinematographer must photograph an 18% gray card, such as the gray card plus. The telecine engineer (or colorist) uses TEC to setup the telecine to a position where the controls are centered and the saturation is at an appropriate level. Using only the primary controls, the operator calibrates each channel of the telecine to the RGB values provided with the TEC film.

The settings are stored in memory to be recalled when exposures on client film are evaluated. The film image of the gray card is read with the setting stored in memory from when the telecine was calibrated according to TEC film. Using a waveform monitor, the colorist reads the RGB voltages or IRE values from the gray portion of the card shot in the scene. Using the appropriate look-up table, the red, green and blue measurements are converted to TP.

Some telecine equipment manufacturers have developed systems to work along with TEC film for automated readout of transfer points. Real time readings as well as computer logging are now available.

The TP Table is used to convert the measured IRE values into transfer points. (There is a separate table for each Kodak Motion Picture film.) A separate set of tables for voltage is also provided. The table has a range of −3 to +3 stops of exposure, which corresponds to approximately 42 TP, where about 7 TP equal one stop in exposure.

TP are related to printer points, but are not equivalent. Each processing laboratory has its own normal printer points. TP are equivalent worldwide if the TEC film is used properly. However, in use, a change of one TP is equivalent to one printer point. The exposure position is quantified removing ambiguous terms such as a little light or a bit heavy which different people can describe to similar conditions. The basis of the exposure reporting is similar to the well accepted printer points used for many years. The TEC system provided a worldwide telecine standard for exposure determination.

SOUND

In the film projectors used in camera type telecines, the sound reproducer is located just below the picture gate. An optical system focuses a slit of light from an exciter lamp onto the film. Usually this lens is adjustable to accommodate films with the emulsion side either towards the light source of the projector or towards the lens. This is necessary only for 16mm format sound; 35mm is always in the same position with the emulsion towards the light.

A system of damping rollers smoothes out the movement of the film over the sound scanning drum following the intermittent frame-by-frame movement of the film in the picture gate. The sound in the film track is located 26 frames ahead (in the direction of film travel) of the corresponding picture frame for 16mm film, and 21 frames ahead for 35mm.

Variations in the light transmitting properties of the sound track modulate the light beam as it passes through the film. The light falls on a photoreceptor, generating an output which, after suitable amplification, becomes the audio signal. The optical system in the sound reproducer must be sharply focused on the emulsion side of the film, where the sound track is located, for best high frequency response. This is a critical adjustment because the focal length of the lens is quite short; a change of only 1/1,000 in. can cause noticeable high frequency loss. For 16mm, the lens usually has only two positions: the back or the front of the film. On 35mm reproducers, the lens is fixed after factory setup.

If the intensity of the scanning beam across the width of the sound track is not uniform, the reproduced sound may be distorted or low in level. Lack of uniformity is usually caused by dirt accumulating in the slit of the optical system. Incorrect positioning of the exciter lamp filament will also adversely affect sound quality. Films that are scratch free and clean will also produce better sound quality.

Continuous film motion in flying spot and CCD type scanners makes sound reproduction much easier, since there is no need for a system of damping rollers at the sound drum as in intermittent projectors used in camera type telecines.

CCD telecines and flying spot telecines scan optical and magnetic sound tracks directly at the capstan. A roller is mounted between the capstan and the picture gate to obtain the necessary picture/sound separation for 16mm and 35mm films.

Many 16mm films in circulation have magnetic sound tracks consisting of a stripe of iron oxide coated on the side of the film facing the projector lamp. The playback head is located 28 frames ahead of the corresponding picture frame. Projectors for television service usually have selectable optical and magnetic sound playback devices.

The azimuth of the magnetic playback head must be properly aligned. Ideally, the gap in the head should be positioned at exactly 90 in the direction of film movement. An incorrect azimuth setting results in loss of high frequency output, as does separation in the head-to-film contact. Periodic cleaning is essential to remove buildup of oxide that prevents good contact.

Today, most film is transferred to videotape for broadcast and it is easier to control the cleanliness, alignment and pristine condition of the telecines' optical systems for both picture and sound in the transfer house.

In post-production and high quality film-to-video transfer operations, 35mm color negative film is the preferred format, instead of making prints from the negatives. The sound is supplied on a separate high quality magnetic film that can be synchronized with the pictures. This way, the best television picture and sound can be obtained for broadcast or for mastering videocassettes for the home VCR market.

Significant quality improvements have been made in recent years in the sound amplifiers and speaker systems of home television sets. Better frequency response, signal-to-noise ratio, and even stereo sound now requires the transmission of clean, quality sound signals. This is possible whether the sound is transferred from film or tape, and if the usual precautions for dirt free, scratch free operations are followed.

FILM HANDLING, CLEANING, STORAGE

In the normal operation of a TV station, it is not easy to ensure that all films are spotlessly clean and free from physical defects. The methods of film handling adopted by station personnel and the degree of tolerance for image imperfections that broadcasters allow sometimes aggravate these conditions. Motion picture film will stand a considerable amount of abuse, but scratches and abrasions can permanently damage a film.

Films can be scratched in any piece of equipment where it is drawn over metal or plastic surfaces. A tiny burr in the gate of a telecine projector can put a scratch on the film from one end to the other. Constant attention is needed to make sure that imperfections are not being caused by faulty or dirty equipment. In this respect, film and tape handling requirements on their respective reproducers are very similar for cleanliness and meticulous maintenance of the mechanical parts.

Abrasions are usually caused by careless handling, such as excessive or erratic rewinding speed. Another very common cause of abrasions is the tightening of a loose wound roll by pulling on the end of the film causing the convolutions to rub on each other. A particularly bad practice is to start winding at high speed and then allowing the film reels to coast. The film should be rewound at constant low tension, sufficient to wind a tight roll without slipping.

With a little practice, the probable cause of a film scratch can be identified by noting the nature of the damage. A straight, uninterrupted scratch parallel to the film edges is most likely caused in a machine where the film moves continuously such as an FSS or CCD telecine. Camera tube telecine projector scratches may show some discontinuities caused by the intermittent action of the claw and gate mechanism. Wave scratches are usually caused by careless cleaning practices.

Torn or damaged perforations are usually caused by faulty projectors or damaged sprockets. Damaged perforations likely will give rise to unsteadiness in the television pictures. (Kodak Publication H-23, *The Book of Film Care* is an authoritative guide on all aspects of film handling, pages 51–83).

Film Cleaning

When film and slides are reproduced in the television system, any particles of dust and dirt as well as scratches and abrasions on the film surfaces are visible in the transmitted images. Electronic enhancement of the images invariably enhance the picture defects as well.

The large network centers rarely broadcast a scratched or dirty film, but local stations often have no choice as they may receive prints that have been used many times. Few stations seem to have the time to clean every film prior to telecast and replacements for defective prints are seldom available.

The preparation of film programs for on-air release or transfer to videotape usually involves much handling and rewinding, often at a high rate of speed. If the film's reels are bent or damaged, the film base and emulsion may be scraped off the edges of the film and trapped in the convolutions of the wound roll.

Static charges on the film surface, generated by high speed handling, can cause the particles of film support, film emulsion or other airborne dust particles to cling to the film surface. Smoking in film handling areas should be forbidden, as ashes and smoke residues can be deposited on floors, work surfaces and eventually on the films. Personnel traffic should be restricted in film cleaning and handling areas as dust and dirt particles emanate from clothing, shoes and even hair. It is not necessary to create an antiseptically clean area to handle film, but clean room practices and common sense can prevent problems in film transfers or transmission.

Cleaning films to remove dust and dirt particles is neither difficult nor time consuming. There are several methods that have been in use for many years. The oldest, simplest and most economical method is the use of *plush pads* and a cleaning fluid. In this method the film is placed on a rewinder and a plush pad is moistened with a cleaning solution, such as pure isopropyl alcohol. The pad is folded over the film strand and the film is slowly pulled through the moist pad and wound on a takeup reel. A fresh section of pad should be used for every few hundred feet of film, depending on the amount of dirt being removed. Care must be taken to keep the pad moist but not so wet that the film strand is wet when wound. If the film surfaces are wet as the roll is being wound up, drying marks may appear on the film surfaces. Cleaning pads can be made from pieces of plush velvet, 12 x 16 in. in size, folded over once and sewn around the sides with the edge turned inward. This is to prevent any plush strands from getting onto the film. These pads may be obtained from photographic supply stores.

Solvent film cleaning machines are available to provide excellent cleaning, proper film handling and high volume. These machines are available from several manufacturers and can be sized for various capacities. Both the ultrasonic cleaners and the plush pad cleaning technique utilize organic solvents. These solvents were especially selected to remove fingerprint grease, wax pencil, and other markings found on film. When used according to the manufacturer's recommendations, properly stored and used with adequate ventilation, they are reasonably safe. New controls for use, storage and disposal of solvents and solvent vapors have made these methods of cleaning more expensive and less desirable.

Film Splicing

Many different methods can be used to prepare films for on-air release or transfer to videotape. For example, a film program and all the short commercials, promos, etc, to fill a full half-hour broadcast period can be spliced together into a single, complete roll ready to run in a telecine projector or scanner.

Another method is to splice together all the commercials and promos on one reel, with short sections of leader between these items, and run the reel on one telecine projector or scanner while the main program is running on another. A complete on-air program or a videotape transfer can be produced by switching from one projector or scanner to the other. Still another method is to transfer the film program and all the commercials, promos, etc, to videotape and package a complete program by electronic editing.

Whatever method is used in a station, some film splicing will always be needed. Each of the splices should be considered as a potential hazard in the telecasting of a program or in its transfer to tape. If a splice comes apart in the telecine, operating schedules will be disrupted, with accompanying losses of time and money and potential damage to the film.

Splicing film is not a difficult or complicated procedure. Good splices that will stand up to normal handling and repeated projections, running backwards or forwards if necessary, can be made in two different ways: either by joining with cement or by applying transparent adhesive tape to the two sections of film at the joint. (See Kodak Publication H-23, *The Book of Film Care*, pages 60–67.)

Cement Splices

Triacetate base films may be spliced using either cement (solvent) splices, or using transparent splicing tape. Polyester (Estar) base films may be spliced with either splicing tape or an ultrasonic (thermal fusion) splicer.

When a cement splice is being made, the ends of the two pieces of film to be joined are overlapped in the splice area and the emulsion is removed from the lower section of film.

Then a little splicing cement is applied to the scraped area, and the base side of the other piece of film is placed over it and pressure is applied until the cement dries.

Film cement is a solvent capable of dissolving the film base. Usually films are spliced in a small hand-operated machine, which has two hinged platens, registering pins to accurately position the two film ends by the perforations, a scraper for removing the emulsion in a narrow strip in the splice area, a knife to cut off the two ends neatly and a hinged pressure block to hold the two ends in contact after the cement has been applied. The film cement is stored in a small bottle, with a tiny brush to apply the cement sparingly in a narrow area of the film at the splice. A good splice has sufficient strength after about 20 seconds to be removed from the splicer and wound up on a reel. Cement (solvent) splicers will *not* work with polyester (Estar) film base.

Tape Slices

Tape splices are easier to make, are less likely to come apart or break during projection than cement splices, and can be made on any film base. The tape used for making splices is clear, with an adhesive backing. Many different types of film splicing tape are available.

Tape splicing machines, readily available, make application of the tape and finishing of the splice very easy. An important advantage of tape splices in a television station is that no frames are lost when two films are being joined together. The film ends are simply butted in the space between frames and the splicing tape applied. Another advantage is that tape splices can be taken apart easily later on, with no damage to the film frames, by simply peeling off the tape.

Storage

Store processed film according to the recommendations in standard ANSI IT9.11-1992. SMPTE Recommended Practice RP 131-1994 also contains recommendations for proper film storage. For short-term *active* storage and projection (telecine rooms and transfer facilities), room temperature of 20 to 25 degrees Celsius (68 to 77°F) at 50 to 60% relative humidity is recommended. Avoid prolonged unconditioned storage at high temperatures or excessive humidity. For medium-term storage, store at 10°C (50°F) or lower, at a relative humidity of 20 to 30%. For extended-term storage (for preservation of material having permanent value), store at 2°C (36°F) or lower, at a relative humidity of 20 to 30%. Enclosed long-term storage (sealed cans) will benefit by the use of Molecular Sieves in the storage container. Molecular Sieves are available from FPC, 6677 Santa Monica Boulevard, Hollywood, CA 90038, Telephone (213) 468-5774.

Additional information on the proper storage and handling of processed film is contained in Kodak publication H-23, *The Book of Film Care*.

FILM AND THE FUTURE

HDTV has been under development for nearly three decades, and although worldwide standards are stil evolving, there is general agreement that HDTV will occur in the near future. HDTV has been used for limited program production in a few studios, and Japan began broadcasting an HDTV signal (MUSE) in 1989. However, there is still significant disagreement over which signal formats are the best choices for production, distribution, and broadcasting.

Film has been, and continues to be, the worldwide standard for the origination and display of theatrical entertainment programs. For over 40 years, it has been the preferred medium for the origination of prime-time television programs.

There have been several studies made of the film and HDTV mediums for video program productions. The performance parameters compared were sensitivity and speed, gray scale transfer functions, contrast range, MTF and limiting resolution, noise and granularity, color reproduction and artifacts. The results of one of the more recent studies was presented in a SMPTE paper by G. Kennel, L. DeMarsh, and J. Norris entitled *A Comparison of Color Negative Films and HDTV Cameras for Television Program Production*. The conclusion reached was that 35mm motion picture film offers significant advantages as a medium for the production of HDTV programming, even if a single HDTV standard is chosen. Sensitivity, scene contrast range, practical and artistic flexibility, speed and worldwide standardization are all advantages of film origination for HDTV.

Film-Video Interface Definitions

The definitions on page 12 are not intended as literal television or motion picture definitions, but are meant to afford a mutual understanding of terms.

APPENDIX A: STANDARDS

The Society of Motion Picutre and Television Engineers (SMPTE) provides continuous development and updating of standards, Recommended Practices (RP), and Engineering Guidelines (EG) for film and television. A complete listing may be obtained from SMPTE.

SMPTE
595 W. Hartsdale Avenue
White Plains, NY 10607-1824
Tel 914-761-1100
Fax 914-761-3115

Motion Picture	Television
Color	Chroma
Overall printer color balance	Chroma Phase
Negative or positive scratches	Dropouts
Scratch removal process	Dropout compensator
Release print	Dub
Edited negative	Master
Dupe negative, internegative	Submaster
Poor printer contact, printer slippage	Banding
Fine grain negative	High band
Projector and screen	Picture monitor
Film printer	Video recording head
Sound recorder	Audio recording head
Printer & process control system (for proper color balance)	Vectroscope
Method used to establish gamma and density	Waveform monitor
Motion picture projector	Film chain
Projector changeover	Multiplexer
A & B printing	A & B mix
Measurement of Density	Video level
Sets D-max	Pedestal
Frameline	Blanking
MP Camera	Video camera
MP film	Video magnetic tape

REFERENCES

1. Eastman Kodak Company, *KODAK Telecine Analysis Film User's Guide*, Pub. H-8-22.
2. Eastman Kodak Company, *KODAK Telecine Exposure Calibration User's Guide*, Pub. H-8-07.
3. Eastman Kodak Company, *Cinematographers Field Guide-Motion Picture Camera Films*, Pub. H-2, 1992.
4. McMurray, "Telecine Techniques to Retain Film Subtleties," *BM/E*, Nov. 1998.
5. Reinking, Franklin, "Film to Tape Mysteries Unraveled," *American Cinematographer*, Sept. 1988.
6. Ryan, Rod, *ed. American Cinematographers Manual*, 7th Edition, 1993.

BIBLIOGRAPHY

1. Bancroft, *Pixels and Halide: A Natural Partnership—Point of View*, 103:306, May, 1994.
2. DeMarsh, LeRoy, *Displays and Colorimetry for Fugure Television*, 103:666, October, 1994.
3. DeMarsh, L., *Evaluation of Color Reproduction in Film and Television*, 95:624, September, 1986.
4. DeMarsh, LeRoy, *Assumptions in Television Colorimetry-Tutorial*, 103:110, February, 1994.
5. Kennel, G., DeMarsh, L., Norris, J., *A Comparison of Color Negative Films and HDTV Cameras for Television Program Production*, 100:337, May, 1991.
6. Kennel, Glenn, *Digital Film Scanning and Recording: The Technology and Practice*, 103:174, March, 1994.
7. Kriss, M. Liang, *Today's Photographic Imaging Technology for Tomorrow's HDTV System*, 92:804, August, 1983.
8. Lees, et al., *High Performance CCD Telecine for HDTV*, 99:837, October, 1990.
9. Lisk, K. Pytlak, J. H. Barrett, *New Tools for Improved Telecine Quality*, 93:6, January, 1984.
10. Poetsch D., *FDL60—An Advanced Film Scanning System* 93:216, March, 1984.
11. Richards, D., and DiGulio, E., *Film to Video Transfers: Time for a Change—Point of View*, 103:85, February, 1994.
12. Snider, D., and Kennel, G., *Digital Moving-Picture Exchange: File Format and Calibration*, 102:712, August 1993.
13. Thorpe, Laurence, *HDTV and Film-Digitization and Extended Dynamic Range*, 101:486, June, 1992.
14. Thorpe, L., Nagumo, F., Ishikawa, K., *A Comparison Between HD Hyper-HAD CCD Camera and Color Film for Television Program Production*, 103:364, June, 1994.

5.15 TELEVISION CAMERA ROBOTICS

ROBERT S. MURCH
RICHARD D. SLENKER, JR.
FOX TELEVISON STATIONS, WASHINGTON, DC
WITH CONTRIBUTIONS FROM
BRUCE LEVY
AND EDMUND A. WILLIAMS

INTRODUCTION

With the application of high speed microprocessors and improvements in servomechanisms, robotic camera systems are now available at an affordable cost and level of sophistication which rival the performance of a competent human camera operator in certain production situations. Performance and reliability have reached a point that a television station or production facility must seriously consider whether the use of camera robots can be an operational as well as financial asset to the facility.

Servomechanisms have been used in industrial settings for many years, from assembling automobiles to assisting in brain surgery. Their precision is well-known and the technology well-developed. It is only recently that robotics has been coupled to studio cameras in order to reduce operator costs, improve shot repeatability and reduce errors. While camera robotics may not be a replacement for camera operators in fast-moving or dramatic productions where a high degree of operator flexibility and creativity is required, they are ideally suited for many less demanding productions.

News programs and productions requiring exact repetition of shots, such as commercials, are perfect applications for camera robotics. Moreover, most productions such as news and talk shows are scripted to a degree where use of camera robots can add to the quality of the presentation due to the smoothness and repeatability of shots.

CONSIDERATIONS FOR USING CAMERA ROBOTICS

An important reason for considering camera robotics is the potential for labor cost savings. Other, operational considerations include the following:

- Repeatability of shots
- Scheduling
- Smooth execution of complex camera work
- Placing cameras in dangerous or normally inaccessible to human operators (point of view cameras)
- Reliability and durability

Labor Costs

The cost of camera robotics varies for different manufacturers and different options. Consider the return on investment of a three camera robotic system that costs $200,000 to purchase and install, for illustrative purposes, and is depreciated over five years. Simply depreciated, the cost of the equipment is approximately $40,000 per year. If the system is operated by an existing video operator and if the station is employing three camera operators at a salary of $500 per week for each operator (or $26,000 per year, per operator) for a total of $78,000 per year, the equipment will pay for itself in about two and one-half years.

Looking at cost trade-offs another way, there is a savings of $78,000 per year in salaries (exclusive of benefits) less the $40,000 cost of capital for the equipment for a net savings of $38,000 per year. The return on investment is further improved with each additional shift of usage plus any savings of employee benefits.

Another payback calculation example: Assume three camera robot systems are purchased at $100,000 each for a total of $300,000. At 10% simple interest and a three year payment plan with equal monthly payments, each payment is $10,833 or $130,000 per year.

Further assume that a station or production facility has 6 full-time camera operators in two shifts at an average of $26,000 each plus 25% overhead for a total of $32,500 each or $195,000 per year. If four of the six operators are replaced (two from each shift) at $32,500 per year each, the savings is $130,000 per year or break-even for the first three years. One camera operator remains on each shift as the robot operator or to be on call for special production shots.

If salary or benefits is higher than those shown above or if the purchase or lease plan is longer than three years, actual cash savings can be realized immediately. A detailed analysis of the station production requirements versus the cost of staff and the costs associated with leasing or purchasing camera robots must be performed before initiating an investigation into replacing skilled studio personnel.

The increase in workload on the maintenance staff to care for the robots, including lubrication, adjustment, repair, testing, software checks and electronic service, must be factored into the decision to use robotics. However, the incremental increase can be expected to be relatively small.

The increase in additional production planning effort may add an extra few percent to the production costs.

This is due to the time needed to setup, rehearse and store the positional information for each shot. However, once done for a given production, less time will be needed for future editions of the same program.

Scheduling

Camera robotics provide a high degree of scheduling flexibility as they are available 24 hours a day and are unaffected by holidays, vacation, illness or overtime.

Repeatability of Shots

Camera robot systems have a degree of precision that provides highly accurate repeatability of all programmed camera shots and camera moves. Once a difficult shot is programmed, the system will execute the shot on-air exactly the same way it did in rehearsal. This factor is important when repeating scenes or to avoid accidentally shooting offset, thus saving countless re-takes or on-air mistakes. Camera robot systems also are designed to accommodate last moment changes and execute them with the same degree of accuracy as pre-programmed shots. From the program director's perspective, the repeatability of shots, and the ability to quickly accommodate last minute changes, provide an additional level of confidence that the production will occur as planned in rehearsal.

The servos that perform the movements of camera robots are able to position the pan-tilt head within 0.1°. The pedestal (dolly and trucking movements) will return to the same place in the studio to within a fraction of an inch and maintain the aim to within 0.1° using the tape guidance system or other guidance techniques. Slightly less accuracy can be obtained when the pedestal is operating in a free mode, that is off the tape or unable to obtain a reference. In any event, the accuracy is equal to or greater than what most human operators could produce.

Movement rates of the pan-tilt heads and pedestals are generally slower than what human operators can perform. This is due to the mechanical effort of the servo needed to move the heavy camera, prompter and pedestal and the desire to keep movements slow enough to prevent injury to someone in the way of the movement.

Studio productions may require dozens of different camera positions and up to 100 different shots by multiple cameras with varying angles, zoom positions and framing differences. Camera robot systems have provisions to store hundreds of shots in memory (for longer, more complex productions or multiple shows) and can interface with zoom lenses, iris controls, camera control units, lighting systems and virtually any other computer controlled studio facilities with external data ports (see section on interfacing).

Smooth Execution of Complex Camera Work

Full featured camera robots will execute virtually all the standard moves of similarly equipped manual cameras including:

- Pan
- Tilt
- Zoom
- Focus
- Dolly (in and out)
- Trucking (left, right)
- Arc (curve left and right)
- Pedestal (up, down)

Each movement can be made at different and varying speeds and have gentle starts and stops.

The manner in which the transition from one shot to another occurs may be critical in some production operations. Some robotics systems allow shots to be linked together along with varying speeds of transition.

Camera robot systems allow an operator to control several cameras at the same time. For more artistic, creative productions a director can plan a number of different, complex, simultaneous camera moves and program them into the system to occur in a particular sequence. When these moves are recalled, they will be replayed exactly as rehearsed, every time they are recalled. For productions which require repetitive shots, such as news programs, this ability relieves the boredom human operators suffer and the subsequent potential for error.

Some robot control systems will also link the zoom lens focal length and focus controls to the camera movements so that as desired objects within the field of view change distance, the focus will track as well, much as a human operator would do. However, it should be noted that, if the mechanical system of the zoom lens is operating unevenly (sticky), the robotics system cannot correct such a defect.

Another caution is that robots will only perform the tasks that have been given to them, exactly as given. If a scene that is changed after rehearsal, a human operator can make an instant and creative decision on adjusting shots while the robot cannot automatically make the on-air adjustment.

Compared with productions with human operators, more advanced planning may be required on the part of production personnel in setting the shots and entering them in the robot's memory. For example, the robot will not automatically focus the lens on a scene but will follow orders to change focus only if the operator programmed it to do so in advance.

Operation of Cameras Located in Inaccessible Places

Advances in robot technology and the use of the smaller, high quality color television cameras make it possible to obtain shots from places and angles never previously thought achievable. Camera robots allow great flexibility for that special point-of-view shots that give the audience fascinating and informative views. Examples of these locations include studio ceilings, on racecars, roofs of sports stadiums, tight spots along parade routes and masts of microwave trucks.

Reliability and Durability

The reliability of any system is dependent upon the reliability of the individual components and subcomponents that make up the system. In the case of camera robotics, the principal components are the servomechanisms and the computer control systems. Servos, by their very nature, are very reliable and if operated at less than their rated loads, will last indefinitely. However, in order to achieve a high level of reliability, a regular program of inspection, cleaning, lubrication and preventative maintenance must be instituted for the mechanical parts of the robot system.

Servos are designed to withstand a certain amount of physical abuse. Some servos have built-in clutches on motors or gear trains to prevent stalling or stripping of the gears. Others can be stalled without burnout. Most will provide an indication of inability to react to commands, such as when an obstruction prevents a pan-tilt head or pedestal from moving to the desired position.

Servomechanisms are designed to operate in relatively hostile environments such as the high heat and humidity of summer and in cold winter conditions. They are, therefore, suitable for use in locations that human operators would find uncomfortable. The range of conditions in which the servos can safely operate is likely to exceed the conditions acceptable for the camera itself.

The computer systems of most camera robots use PC components. The normal lifetime of most of these parts is approximately several years. Little maintenance is required other than the usual routine inspection and cleaning. However, if mechanical memory devices are used (hard and floppy disk drives), an occasional check should be made of the disks using standard software available for this task.

Because most camera robots are likely to be operated 16 to 20 hours a day, and may in fact be left on 24 hours a day, the disk drives are the most likely parts of the computer to fail. A backup drive or computer is recommended to help bridge the time when such a failure occurs.

INSTALLING CAMERA ROBOTICS

The major installation issues to be addressed are:

- The impact installing robots will have on the staff
- Potential locations for camera robots
- Smoothness of the studio floor
- Location of the robot controls.

Personnel Considerations

The installation of camera robots can result in more effective uses of production personnel. How management deals with the staff, who may believe that robots will replace human operators, may determine the level of success of the installation more than any other factor. Depending upon individual situations it may be possible for reassignment or attrition to help adjust staffing arrangements.

In addition to the changes in the technical and production staff due to the installation of robots, other departments in the station must also be directly involved in the project.

The use of camera robots will affect how a production is designed and executed. All production personnel, including producers, directors and other production managers throughout the facility must be aware of the advantages and the limitations of using robots. Initially, this awareness could begin with the Engineering Director providing information about robots and how they can improve production and reduce production costs. Later, after the staff has become used to the idea of using robots, production staff can be trained on how best to use the camera robots. After the initial introduction of the robotic system, it is essential for engineering and production departments to work jointly to develop daily standard operating procedures for the new robots.

Potential Locations for Camera Robots

While the studio is the obvious first choice for using robotic cameras, it is by no means the only location. Newsrooms are popular sites for robots where the camera is not required to be as full featured as in the studio. Newsroom cameras may not need to be mobile and simpler robots (equipped only with pan, tilt, zoom and limited side-to-side travel) may be all that is required. In any event, when considering installation of camera robots, carefully study how the camera will used. Consult with other stations that have installed robots and review the performance limitations of the system under consideration to insure that it meets the needs of the facility. While it is relatively easy to have an operator move a camera to a more suitable location, moving the robot may not be possible without considerable time and expense.

Camera robots also can serve to relieve operator boredom in production situations where a camera must be left in place for long periods and used only occasionally. Examples include parade routes, city councils or other public forums, racetracks, elections and telethons. All that is needed for control is a data circuit between the control point and the robot. The operator at the control position can observe the response of the robot by the action shown on the monitor for that camera.

Robots come in a wide variety of configurations and cost ranges. Some of the simplest devices are the older, yet still desirable, pan-tilt mounts. At the next level are those with memory for different shots and variable speed, followed by those with capability for movement along one or more axes. The full featured robot adds elevation to the other movements and some have scissor type mounts (or cranes) to extend the elevation range at both upper and lower extremes.

Studio Floors

Camera robots weigh several hundred pounds. Occasionally, they and may be required to execute shots while in motion. Thus, the surface of the studio floor is extremely important to the on-air look the cameras will produce. One of the best studio floors is built of concrete with a smooth epoxy paint covering that, together, provide a degree of levelness of less than

1/8 in. height variation for every 12 linear ft. Many studio floors have a concrete base but have been covered with tiles that may not be entirely suitable for the weight of robots or may be chipped and require replacement.

Another possible and popular location for camera robots is the station newsroom. Many newsrooms have computer style raised deck flooring that must be evaluated for suitability. The manufacturer should be consulted to insure that the floor will be suitable for the robots.

Control Point

An important part of installing camera robots is deciding where to locate the control point of the system. Possibilities include the studio, production control room, camera control position or a dedicated control room. An advantage of having the robot controls at the camera video control position is that the video operator can combine this function with that of the robot controls.

One approach is to have the camera control or video operator also assume the duties to control the robots, either from a studio control room or a dedicated video control room. Some installations locate the robot operator on the studio floor. In some instances the control position is elevated. This has the advantage of giving the operator direct view of all camera movements.

Some systems have cable length limits between the camera control heads, the robot control panels and the rack mounted control processing and memory storage units. Such restrictions may limit the operation for placement of control panels or rack systems.

CAMERA ROBOT NAVIGATION SCHEMES

Camera robots use sophisticated computer techniques and servomechanism technology to provide the quickness, smoothness and repeatability required to be effective. In addition to robot systems the pedestals are manually moved to a location by a person on the studio floor, camera robot pedestals that can automatically move about the studio are also available. A variety of schemes are used by the different manufacturers for the system to know where each pedestal is located. The sophistication of these navigation systems is such that most systems will not allow cameras to run into each other. As discussed in the following section on Safety, there are mechanisms on the camera pedestals, such as proximity and sonar detectors, to protect the cameras if they encounter an object.

Studio camera robots must have some means for determining their position in the studio and with respect to other robots. There are several different approaches to determining how the robots move about the studio which include mechanical tracks, tape tracks, targets, and wall mounted bar coded signs. Of paramount concern is safety to both studio personnel and other robots and facilities (monitors, set pieces, microphone and lighting stands) within the range of the robots.

New studio procedures must be developed and established when employing camera robots. Camera robots depend upon reference marks which cannot be covered or damaged. The robots also may move without apparent warning during setup, rehearsal and on-air production. Camera robots do not have peripheral vision! Therefore, a new awareness of studio activities is required of all production personnel. Moreover, the damage a robot can cause to monitors, lenses and other studio equipment may cost more than the operator that was replaced.

There are several methods and associated technology that are employed for controlling the position and keeping track of camera robots.

Mechanical Track System

The oldest and most limited method is the X-Y rail system which requires the construction of wooden or metal tracks or rails on the studio floor on which the robots travel. This system was adopted from the camera dolly systems used in motion picture studios. While the technique provides extremely smooth and repeatable shots, substantial work is required to move or reconfigure the tracks if the studio set is changed or if a different perspective is desired. A simple track system may be ideal for use in the newsroom or dedicated news studio where limited movement is all that is needed.

While movement of the robots is limited by the rail system, it is important that a sufficient number of proximity and contact sensors be included to prevent cameras from colliding, cables becoming cut or damaged or personnel being injured by the moving equipment.

Tape Track System

This basic navigation system uses a metal foil tape placed on the studio floor that functions as a track for the pedestal to follow. Some robot control systems allow the pedestal to move away from the tape for short distances. The robot then relies on counting wheel revolutions (tachometer pulses) to compute its position. Disadvantages to this system are the need to change the tape whenever there is a change in set placement or damage to the tape caused by moving heavy objects over it.

The Target System

The tape track technique can be expanded and simplified by having a home target on the studio floor. Each camera robot uses its target to establish a reference just before a production and is then free to roam the studio floor. Of course, the further the robot moves from the target the greater the accumulated error will be, especially after changing positions several times during a production. Just as the tape track must be protected from damage, the target must also be protected yet made available to the robot whenever it is necessary to re-establish its reference.

Bar-Coded Signs

Another robot navigation approach is to attach bar code signs on one wall of the studio. Each camera robot pedestal then uses a laser in the pedestal base to scan the bar codes on the wall. By knowing the location of the bar code cards, the system can calculate the location of the pedestal. While this technique employs substantial computer programming power, accidentally blocking the view of the signs can render the robot blind to finding the reference. The operator will be signaled when this happens, but the carefully designed production may be adversely affected as a result.

Each of these navigation schemes has its advantages and disadvantages. When selecting a camera robotic system, the user should evaluate each scheme to decide which one is best for the facility in which the robots will be working. An evolutionary approach may be employed in which initially a system would be installed with manually positioned pedestals that would later be upgraded to full mobility, perhaps one camera at a time, for economic reasons.

SAFETY

Safety is an extremely important matter with robotics cameras. The cameras must not run into people, the set, or each other! Several methods are employed by different manufacturers to provide various levels of safety.

- Limits: Virtually all robotics servos are current or torque limited. This means that if, for example, the camera pans left and the lens pushes into a floor monitor, the servo will sense the obstruction and stop movement in that direction. In some systems the servos will back up a little. This limiting action can apply to any of the several directions of movement of the robot. Moving pedestals may have additional safety features. For example, infrared sensors or microswitches on the base of the pedestal will sense any obstruction and stop the pedestal movement.
- Observation: One way of observing the movements of the camera robots is to install a ceiling mounted camera that takes a wide shot of the studio. The output of this camera can be displayed on a monitor in the camera robotics control area.
- Signs: The installation of large and bright signs on the sides of the camera robots and the studio doors and walls will serve to remind production personnel that the cameras move by themselves and without warning.

WARNING!
KEEP CLEAR!
CAMERA ROBOT MAY MOVE
WITHOUT WARNING!

- Cables: Making sure that the cameras do not get tangled in their own cables is both a safety as well as an operational consideration. Problems of this nature can be alleviated by careful planning of camera movement. The camera cable and the robotics control cable can be sleeved in a nylon webbing. This webbing combined with low skirts on the camera, will minimize cable management problems.

OPERATING CONTROL SYSTEMS

Three of the most common types of controls for robotics cameras include: joysticks, data tablets, and touch screens. Most control systems can operate several cameras (four to eight) at once. The software in most control systems is designed to prevent the operator from requesting the cameras to do something that cannot be done (move to the other side of the studio in 5 seconds); is dangerous (move to the same location as another camera); change position while on-the-air (unless the operator over-rides the warning); or exceed a built-in limit or range.

Joysticks

This popular and relatively simple method provides simple direct control over the camera movement. Typically, one camera is selected at a time for direct control in the system. Using the joystick, the operator sets up the shot to the director's satisfaction. The rate of movement is also established. The shot is then stored as a number or an alphanumeric designator in the computer memory. The shot can later be recalled during rehearsal or broadcast using that number or alphanumeric.

Data Tablets

These devices are similar to those used on some video graphics generators. The operator can draw a studio layout on the tablet, indicate the positions of the camera and then teach the system the desired camera shots. Shots can then be recalled by touching a specific area on the tablet. For example, touching the tablet at a location representing the front of the sports desk could move a camera to a wide shot of the sports desk. Touching the tablet at a point behind the desk could cause the camera to go to an effects shot. In addition to positional information the system also stores zoom lens settings and pan-tilt positions.

Touch Screens

Similar to other touch screen controlled devices, such as graphics generators, computers and test equipment, this control system allows the operator to set and recall camera cues by the touch of a finger on a screen. Typically, these systems use menu driven color touch screens. On some systems on-air tally indication can be provided on-screen that are driven by the tally signals from the switcher. This feature reduces the chances for error by the operator accidentally moving an on-air camera to its next shot.

Options

When the primary usage of a camera robot system is the presentation of news, two other newly developed operations may be of interest: *auto focus* and *auto framing*. Auto focus works on the same principle as an auto focus consumer film camera. The time delay between the transmission and reflected reception of a high frequency sound is used to determine distance that sets the zoom lens focal length. An override control is needed with this option because auto focus may be fooled into focussing on a large foreground object rather than the intended subject.

Auto framing is an option that uses sophisticated comparison technology to adjust the framing of a shot when a moving subject is involved. A predetermined shot is programmed into a video frame store. The system continuously compares the stored shot to the live shot and automatically pans or tilts the camera as required to keep the object and stored frame coincident. A news talent who is prone to body movement on the air can be kept properly framed automatically.

Limits

Most systems, regardless of the type of control, can control multiple camera moves simultaneously. Some systems can also warn of impossible shots. For example, giving a command to move a camera from one end of the studio to the other in three seconds will cause some systems to display a warning message. Using camera robotic systems simultaneously to move multiple cameras, however, can result in added finesse to productions since more complicated moves can be smoothly executed on a consistent basis.

Backup Systems

Because the entire studio operation may be contingent upon the proper operation of the camera robots, it is essential that backup power and computer arrangements be made. In event of an electrical circuit outage that could take the robot controls down, some kind of uninterruptible power supply (UPS) on the computer control system (not the robot mechanisms) to bridge the gap between outage and restoration of line power should be considered.

During installation of the robot system some of the power supplies for zoom and other functions on the camera may be replaced with power from the robot system. To avoid a complete failure of the camera if the robot fails, the original power supplies could be left in place for manual reconnection if needed. This arrangement may also aid maintenance of the system.

The computer control itself is also a source of single-point failure that can take out the entire system. A backup computer would reduce the time to restore operation if the main computer had to be repaired. Of course, a means for switching to a manual control panel could also help solve this problem.

Backup systems become more important if the camera robot controls are interfaced with other studio facilities. More than just a camera position is at stake if a control system failure should occur.

INTERFACE WITH OTHER SYSTEMS

Camera robot systems have varying levels of interface capability. Some provide for extra extension panels (pushbuttons, keyboards, joysticks) to be used for addressing external devices or systems but tied to the robot control system. The cues and control signal can be initiated as part of the shot list for the cameras. Some of the possible options are described below.

Some caution is needed when considering the interconnection of multiple devices employing microprocessors. The interfacing software must be developed, rigorously rehearsed, and carefully documented to avoid one system causing problems with another. The knowledge of how the systems interact must be shared with others so that when problems occur, personnel are available to solve them quickly. Few situations are more exasperating than when it is impossible to accomplish a function because the computer is down or a bug has been encountered in the software.

Camera Control Unit

Many robot control systems provide an interface with the camera control unit (CCU) such as iris, pedestal or blanking or black level, color balance, and other controls) if the CCU has a serial data port that can be accessed.

Switcher Automation Systems

The robot controls can be interfaced with some production video switcher and effects systems (if an external serial data port is available) that can, give instructions to the switcher upon selection of a camera move or activate a camera movement upon selection of the camera by the switcher. The result is simultaneous selection and start of movement by the camera.

Lighting Systems

The interface of studio lighting control systems with the robot control system provides the ability to cue lighting sequences along with the activation of camera moves. The camera control could be used to provide a trigger to the lighting system to initiate a lighting change sequence or merely to turn on or off the lights for a set about to be used on camera or a set that no longer is needed. Again, a serial data port is needed on the lighting control system for the interface.

Audio Control Systems

Just as audio-follow-video (AFV) interconnection systems are common in many television stations to switch microphones along with camera takes, similar interconnect arrangements can be made with robot control systems. However, the use of a more intelligent controller than that in the switcher provides the ability to control the audio with more finesse. For example, in situations where a complex video switcher and effects system does not provide reliable tally or AFV signals, the robot controller may provide the ability to turn on a microphone slightly in advance of a camera take to that on-air talent and turn it off some time after the camera is off the talent (to avoid clipping dialogue).

Newsroom Facilities

An interface with electronic newsroom facilities and the robot controller provides the ability to cue prompters, turn on microphones and lights, and activate tallies and other signals. Cues in the prompting text can be used to switch cameras or select pre-programmed shots.

Data Modems

Because the robot control is a computer, virtually any device that can be connected to a modem can be controlled in some fashion. Thus, it becomes possible to control a camera or some other device at some remote location via telephone lines or other data link. For example, cameras and other devices can be controlled via modems at satellite news bureaus or other remote sites.

OPERATIONAL CONSIDERATIONS

Frequently camera robots are used daily for scripted or fairly predictable shows such as newscasts or talk shows. Occasionally it may be desired to use camera robots for telethons or other more complex or spontaneous productions that would be difficult to do with robotics control. On these occasions it may be desirable to have camera operators. Some camera robot systems have switches that allow the robotics to release control of the camera pan head and the pedestal and thus allow for almost instant manual operation. This feature (also called *backdrive*) can also be used as a manual backup in the event of failure of the robotics system. Another advantage of having this feature is to mix manned and unmanned cameras to suit the needs of the production.

Some robots will allow a switch to manual control, memorize the new position, and, after returning to remote control, adapt to the new position and continue with the shot sequence. In other words, on some systems, remote and local control may be mixed at any time.

Operator Training

The operation of robotics camera systems is fairly straightforward. Therefore, training can usually be accomplished in a few sessions with a good instructor. After the operator is familiar with the system, it is important that directors and others involved in studio production agree on a uniform set of commands for robotics camera operation. Rehearsals before the first live session can also be helpful. As with other operating positions in a station, one primary and several backup camera robot operators should be trained for this essential position.

Maintenance

Camera robot systems are generally considered to be reliable and require little routine maintenance other than the general cleaning, lubrication, routine testing, and periodic software and electrical checks that any electromechanical device in daily use would require. Support from manufacturers varies, however, and the availability of parts and service must be resolved at time of purchase. Good documentation on mechanical components, electrical systems, and computer software will help insure fast repairs and fullest use of the system's capabilities.

Some robot controllers use standard personal computer (PC) hardware plus specialized keyboards, tablets, or touch screens to help keep maintenance simple and straightforward.

SUMMARY

Camera robotic systems can be a cost-effective investment for a broadcast television station or production facility. The level of sophistication in computer control and servomechanisms has reached the point where camera robots can provide respectable performance in certain production situations. In addition to the cost savings in operators there is the potential for linking the robots to other automation systems to integrate programming activities. However, consideration of the use of robots must be made based on evaluating the capabilities of the technology that can improve operational efficiencies without causing any degradation of production values.

Careful planning, both of people and equipment, selection of a manufacturer, proper installation, a good maintenance program and a control system with the degree of sophistication that meets the needs of the station can result in an effective, efficient and cost-saving installation.

5.16
LIGHTING FOR TELEVISION

BILL MARSHALL
CARIBINER INTERNATIONAL, INC., NEW YORK, NY

INTRODUCTION

The role of the television lighting director has been transformed due to the continuing development of more sensitive television cameras. The position is now more commonly referred to as the lighting designer, and has become an essential part of the creative team in television production. Lighting techniques in television have become as deliberate and carefully crafted as in film production. Many of the compromises of real time multicamera production are no longer necessary due to lower light level requirements, sophisticated control systems and enhanced editing capabilities. Despite the advances in technology, the lighting designer must remain aware that the home receiver, with its limited range of contrast (30:1 approximate), continues to be the weakest element of the television system.

Television is a two dimensional medium and therefore lighting of the subject is critical to the suggestion of three dimensional forms. Even the impressive resolution of high definition television can only replicate an image as it is revealed by light and shadow. Good lighting will both model the form of the subject and its surrounding and maintain its relative balance of intensity to the rest of the image. Lighting design can be most accurately defined as *constructive* use of controlled light for a predetermined objective.

A major portion of the shots used in television are close-ups; therefore, lighting of faces is the single most important task even though modern robotic lights are often used as a major background element. The standard techniques for modeling faces with light as used in still photography were simply a modern adaptation of the same approach used by portrait painters for centuries.

The basic arrangement of lights for television lighting is commonly referred to as the *two-* or *three-point system*. The first element or *point* is the *key light* which is the principal illumination of the subject's face. It is placed on an axis or slightly to one side of the camera and generally at an elevation of about 30° above the horizontal line between the subject's face and the camera lens. The elevation should be adjusted relative to the particular structural elements of each subject's face (such as length of nose or depth of eyes relative to the brow). The location in practical terms will obviously be determined by where or how the fixture can be supported. In studio work, this means the key light is very often suspended much lower than the general grid height to gain precise positioning.

The second point or component of this lighting approach is *backlight*. Located above and to the rear of the subject, the backlight creates a glow on the hair and a highlight on shoulders separating the subject from the background. The combination of a camera mounted key light and stand mounted backlight is the most prevalent arrangement used in electronic news gathering (ENG) and electronic field production (EFP). Ironically, the sole use of the camera mounted key light often creates a somewhat less than ideal picture, but has come to have direct association with realism.

The third element of the classic three-point lighting system is *fill light*. A single key light can often provide modeling that is too severe in the absence of other ambient light sources. Fill light generally softens and blends the back and key light accents while maintaining their purpose to highlight and separate the form.

While a setup such as described here is relatively easy with one camera and one subject, the reality of television production is that there is generally more than one camera and multiple subjects. To light an actual production, the lighting director will expand the basics of the three-point system to cover all the subjects and for all the camera angles. This often requires clever fixture arrangements wherein the function of a fixture has multiple uses. For example, in a typical two person interview, the lighting design might be arranged such that one subject's key is simultaneously serving as the other subject's backlight.

Simultaneous with careful positioning of light fixtures for modeling, the lighting director must also carefully consider the control of the shadow the light sources will create. Microphone boom shadows can be difficult to control. For each camera angle, the key light and backlight relationship must continue as well as control of the shadow they create. Control of boom shadows can require careful coordination between the boom operator, the director, lighting director and scenic designer; however, the use of wireless mics have greatly simplified this issue. Other common shadow problems, which are generally within the lighting director's control, are elongated nose and chin shadows which occur from improperly positioned key or backlight fixtures.

It should be noted that projected shadows, shapes or patterns can be some of the most powerful techniques available to a lighting director in suggesting time, place, mood or even pure background decoration. The character and quality of light is often more apparent from the shadows it casts. Shadows can be hard, soft, transparent, or even a different color than the apparent source. Because shadows are so important, major lighting arrangements should be carefully planned and drawn to scale. Section views in scale will help to predetermine where the shadows will fall, if there will be any scenic conflicts and most of all, this preplanning will save a tremendous amount of time in the studio or on location.

Once the lighting of the faces is established, the lighting director plans the background lighting. While a TV program can occur in almost any conceivable background, the lighting of faces continues to determine the relative brightness of the backgrounds. The reflective value of a face ranges between 28 to 41% while a white wall can be 96% reflective. A lighting director must therefore be very careful to use lighting to maintain a proper balance since the viewers' attention will be naturally drawn to the brightest spot of a television picture. Lighting intensity along with scenic element selection is critical to controlling the viewers' focus.

The cyclorama (cyc) is a generic background found only in studio lighting. A cyc can easily be over lit and then in contrast over power and appear to darken the skin tones. Projection of light patterns on the cyclorama is often an inexpensive way to create a setting with light, but again, the patterns which decorate a long shot should not appear undesirable or too bright relative to natural skin tones in the close-ups. Elaborate scenic treatment may require all types of built-in specialty lighting and a whole assortment of fixtures for each unique requirement.

LIGHTING LEVELS

There is no one proper level of light, except as required to allow a particular camera to make a good picture under specific conditions. The level required will vary by camera type, lens, type of action, quality of teleprompter glass, existing ambient lighting conditions and other purely aesthetic considerations. The lighting director must control the relative balance of the lighting for every given situation whether it is in brilliant sunlight or in a dark dramatic stage setting. There are often situations where the natural conditions greatly exceed the limited 30:1 contrast range of the television broadcast receiver. Therefore, it is also the lighting designer's responsibility to lead the director and design team away from these difficult conditions.

As with any camera lens system, the amount of light required is a direct relationship between the pickup medium, lens aperture and depth of field desired. Long focal length lenses used in sports events have smaller apertures by design. Since the fast moving action demands a high depth of field, the lighting level must therefore be high enough to satisfy this combination of factors. Conversely, the exact same camera in a dimly lit studio close-up may require a fraction of the illumination level of the sporting event. Good picture quality and proper level can only be determined through the lens of the video camera.

There are many devices available to the lighting director to control the levels of illumination including a wide variety of television studio and location lights. The distance between the fixture and the subject will directly affect intensity. Intensity can also be controlled by scrim, which is actually a wire screen inserted in front of the lamp. The greater the density of screening, the more the light level is reduced. Actually, any medium placed in front of a fixture will reduce its output to a predetermined degree. A wide variety of available sheet diffusion materials will reduce intensity and reshape the pattern and quality of the light. Diffusion materials soften the light and shadows produced. Even colored filters reduce light output relative to the coefficient of transmission of the particular color. For example, a blue filter passes very little light while a yellow-green filter will reduce the output light level very little. The most precise and convenient level control of incandescent lighting is dimming. Compact portable dimmers are also used outside a studio or theatrical environment, but do not necessarily eliminate mechanical means of level control. While dimming does change the color temperature of the lamp, it can still be used effectively.

COLOR TEMPERATURE

For accurate color rendition, the television system must operate within a consistent range of *color temperature*. Color temperature is measured in degrees Kelvin (°K) and varies dramatically from type of artificial light to time of day and atmospheric conditions of daylight. A television camera must be white balanced for the same lighting conditions as the scene will be shot. Segments shot separately which are to be seamlessly edited together must be shot under consistent color temperature light for accurate color rendition between segments. In our daily routine, the eyes experience the effects of different color temperatures while the brain interpolates color rendition. Recalibration of the camera for varying lighting conditions has somewhat the same effect yet it is often important to retain some of the original lighting color and character. The television system is actually quite flexible, permitting color temperature swings of ±300°K without major effect to visible color rendition.

Exterior windows generally present the greatest color temperature problem because exterior daylight is a tremendously different color temperature and is extremely bright relative to the artificial interior lighting. Large sheet filters which correct both color temperature and intensity are available for this purpose; however, if the window area is extremely large and daylight dominates the scene, it is generally easier to work with high color temperature light sources rather than incandescent.

LIGHTING FIXTURES

A designer's choice of lighting equipment differs considerably between the studio and on location. On location, equipment is chosen for its light weight and efficiency. The lighting designer on location must cope with varying natural light conditions from full sunlight to little or no light. Studios conversely are designed and equipped to give the lighting designer absolute control over all aspects of lighting: intensity, color and placement.

Location Lighting

Since lighting design is deliberate control of light for a desired effect, good lighting design may not even require the use of any fixtures at all. In many daylight situations, a lighting director may choose to simply use reflectors and scrims to achieve the desired effect. Movement of the sun is predictable so the survey of the site must be made at the same time of day of the event shoot in order to select workable camera positions and to determine the proper equipment. The sun as a key source is very harsh and requires intense fill light for the talent. Reflectors are very effective for this purpose. They require no electrical power, but must be attended by a person at all times to keep them oriented correctly to the moving sun. Lightweight folding reflectors are convenient to pack and setup but do not present the necessary hard, stable surface needed to provide a smooth field of light. The heavy solid panels are less convenient but are stable. Another way the harsh light of the sun can be controlled in intensity is by using large silks or nets to soften or to shadow the talent. Nets and silks under 6 ft × 6 ft attached to frames are called *butterflies* up to 20 ft × 20 ft are called *overheads*. Wind is an obvious factor in the decision to use these tools.

For smaller remotes, the lighting director's selection of fixtures is often governed by weight and portability relative to lumen output. There is a tremendous variety of extremely ingenuous and lightweight fixtures available. These generally are packaged in kits with a full assortment of stands and grip hardware to accommodate any situation which may occur while on location (see Figure 5.16-1). Many of these kits are designed to incorporate dichroic filters to match the 3200°K incandescent lighting to the 5600°K daylight. This filter reduces the light output quite noticeably.

For larger location work involving daylight conditions HMI® lighting has become the standard of both the film and television industries. HMI® is a registered trademark of OSRAM, although it is often used (or misused) in referring to lamps of other manufacturers. The name refers to the basic elements, which are combined in the lamp's quartz envelope to create the unique color temperature: "H" for mercury (Hg), "M" for the various metal halide rare earths, and "I" for the halogen iodine and bromine.

The highly efficient ballasted arc source lamps come in a wide variety of sizes from 200 to 20,000 W. They provide about 80+ lumens-per-watt at 5500°K (which is 4 times the output of incandescent light). HMI® lighting is often the best solution for interior situations with tremendous window exposure. As in any exterior situation, the lighting director must carefully monitor the changing color temperatures of daylight and remember that the light of a clear sky late afternoon is quite a different color and character than mid-morning when the sky was overcast. Figure 5.16-2 shows a typical HMI® device.

Figure 5.16-1. For remote lighting, compact kits which contain a large assortment of fixtures, stands, reflectors and grip hardware are available. *(Courtesy of Lowel-Light Manufacturing, Inc.)*

Often a lighting director has to shoot in situations where the existing lighting is a low ceiling covered with fluorescent fixtures and they must be used because they characterize the space. Common architectural fluorescent light does not provide a full spectrum light and the most common type cool white is very short of red energy. While it is possible to color correct each fluorescent lamp using a minus green color correction gel it is often impractical. By color balancing the camera under a representative mix of the fluorescent and the 3200°K talent's lights acceptable skin tone can be produced while the background may still appear greenish. Today, there are a wide variety of fluorescent lamps available and some offer quite a high color rendering index. In fact, a line of fluorescent studio fixtures that utilize compact fluorescent lamps have

Figure 5.16-2. HMI lamps are most often used in fresnel lens housings and are available in a wide variety of sizes. The associated lamp ballast is a separate unit from the lamp housing. *(Courtesy of Arriflex Corporation.)*

become quite popular. They are very energy efficient and are an excellent choice under certain conditions.

Exteriors at night for TV require great simplicity to look real. Large HMI's® are very useful because one large source can supply the basic illumination for a very wide area. Within an exterior scene there are always elements supposedly lit by artificial sources either seen or imagined. With the camera balanced for 3200°, quartz lights can be used for the people areas while the overall HMI's® at 5500° will seem very blue. If the HMI® light is too blue, it can be corrected slightly until it seems to be the proper gray-blue of moonlight. Using smaller HMI's as back and rim light will further enhance the moonlight effect. In attempting to provide realism, lighting must suggest the correct mood. Lighting is essential in creating the feeling or mood the production team envisions.

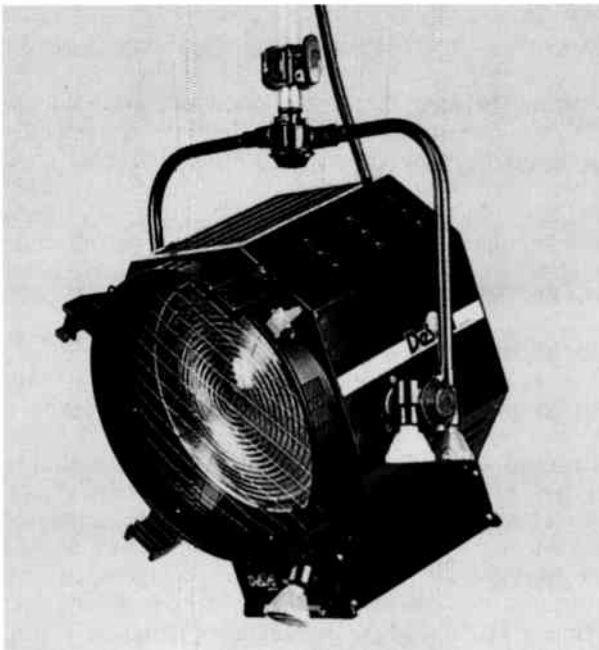


Figure 5.16-3. Studio quality fresnel spotlights are much more durable and have superior optical systems compared to more inexpensive theatrical units. *(Courtesy of DeSisti Lighting)*

Studio Lighting

In a studio the television lighting director has the greatest control over lighting quality. A TV studio is by definition an idealized environment for production of television programming.

In a well equipped studio the television lighting director has quite a variety of lighting fixtures or luminaries available to accomplish the basic objectives of TV lighting: separation, modeling, accent, illumination and directing viewer attention (see Figure 5.16-3). This variety of fixtures and the various qualities of light which they produce have been developed to enhance the efficiency of studio operation. The careful manipulation of light quality by external diffusers, cutters, reflectors, etc., for each shot is still not as common in video production as film but has increased greatly. A complement of fixtures with given characteristics plus the mechanical light control techniques of the film industry provide a television lighting designer with a versatile palette of choices to design the lighting.

Spot Lighting—Fresnels

In a studio, the most common and useful fixture is the *fresnel lens spotlight* (see Figure 5.16-4). The lighting from a fresnel is very controllable and has a smooth even field. By a simple adjustment of the position of the lamp, the fixture will produce either a narrow spot or wide flood beam of light. Equipped with barndoors, this beam can be further shaped to virtually any pattern. Slots in front of the lens allow color, diffusion media or screens for intensity control to be slipped in place. A complement of single and double thickness. and half and full frame screens

Figure 5.16-4. Studio quality fresnel spotlights must offer good quality long leaf barndoors as an essential accessory. *(Courtesy of Strand.)*

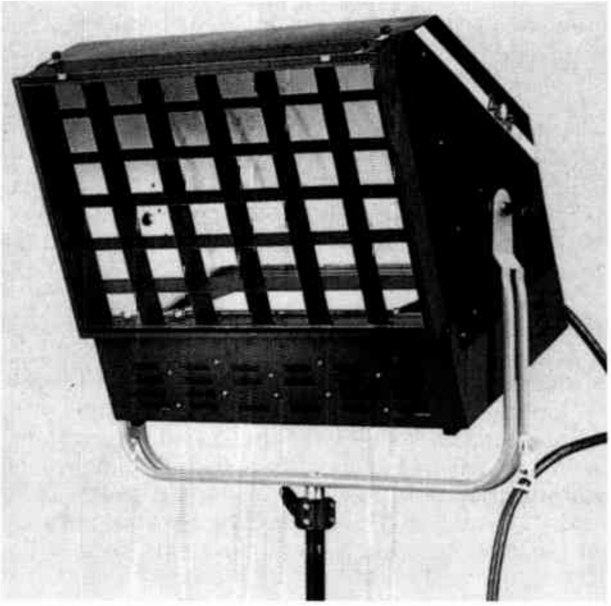


Figure 5.16-5. Spotlights fitted with eggcrate louvers offer controlled smooth even lighting.

should be provided for all fresnels even when dimmers are available.

For most 14 ft to 18 ft high grid studios the 6 in., 8 in. and 10 in. in wattages from 500 to 2,000 are the most common units. In smaller studios or tight applications smaller fixtures are often employed. These units enclose the same wattage lamps in small housings. A number of people prefer these baby size units for their easier handling and somewhat different optical characteristics. As noted previously, the higher the grid, the higher the required wattage fixtures. Therefore, in higher studios, the standard complement of fixtures might be 5,000 and 2,000 W units or even up to 10,000 W. Conversely, fresnel lens units are also available in sizes down to a 2 in. lens with a 100 W lamp.

In selecting a line of fresnels the studio lighting designers must carefully weigh both the optical and mechanical features including:

- *Stability.* Poorly made fixtures will not focus properly once they are hot
- *Balance.* Fixtures should also be well balanced, even with barndoors to remain in focus
- *Accessories.* The manufacturer should also provide a complete line of suitable accessories including barndoor, with leaves that are long enough, screen sets, diffusion frames and stand mounting hardware.

Soft Lighting—Softlights

The softlights are another common studio fixture (see Figures 5.16-5 and 5.16-6). They provide a controllable diffuse base or fill light. In this fixture the light sources are totally concealed and the light is

Figure 5.16-6. Softlights produce a very diffuse source and have switches to select as many lamps as necessary. *(Courtesy of Mole-Richardson.)*

reflected in an indirect manner. These units are not particularly energy efficient, but are unequaled in providing a shadow free light, which can be controlled with eggcrate attachments. They are available in a variety of sizes. Large 4,000–8,000 W units are pretty unwieldy for many applications, but the smaller units in the 1,000 to 2,000 W range are becoming very popular as fill lights, because they are easy on the talent's eyes. With lower light level requirements softlights have become the preferred broad baselight source.

Area Lighting

Scoop floodlights were once very commonly used in studios to build up high levels of base light that is no longer required. Scoops are used invariably with diffusion media, so it is essential that a diffusion frame is ordered. While the majority of lighting can be done with fresnels and some type of diffuse source, there are a vast variety of fixtures which each produce a different quality of light for special applications. For a harder rectangular field of fill light, a broad can be used. It is very similar to a scoop in concept, except its housing/reflector creates a somewhat less diffuse light. It is always used with diffusion media. These fixtures are now most commonly limited to lighting backgrounds.

Other Fixtures

Probably, the most efficient incandescent fixture used in television lighting is the *Par unit* (see Figure 5.16-7). The lamp contains both the reflector and the

Figure 5.16-7. The source 4 Par provides similar output as standard Par 64, at a lower wattage. It has interchangeable lenses for a variety of beam spreads.

Figure 5.16-8. The source 4 ellipsoidal produces extremely sharp pattern for many interesting effects.

lens, which is available in 5 different beam spreads. While limited to very large studios and remotes, the Par is an inexpensive and powerful tool for the lighting director. A new variation of this fixture has been recently developed which provides a similar quality and quantity of light with a lower wattage lamp. It has the added convenience of interchangeable glass spread lens, which match the different lenses of standard Par lamps.

The *ellipsoidal reflector* spotlight is a unique effect light in a studio (see Figure 5.16-8). The optical system of this unit allows the beam of light to be hard or soft-edged and the projected pattern of the beam can be precisely shaped by internal shutters. It is most commonly used to project patterns on the cyclorama. The modern ellipsoidal spotlights with dicroic reflectors and sophisticated optics offer very bright clear projections of patterns, which can be made from photographic half tone-etched metal or lithographed glass slides. These sophisticated patterns are available in a variety of standard and custom designs. Ellipsoidal spots come in a variety of beam spreads but the wide-angle short-throw sizes are most useful in small and medium studios.

Robotic Fixtures

Robotic fixtures have become common place in in theatrical and television lighting. These lights fall into two categories. The first group can position its light by means of a motorized yoke which therefore requires a sturdy mounting position that does not sway from the inertia of the moving unit. The second group utilizes a motorized mirror that reflects the output beam of light into the correct position. Within these two mechanical solutions there are *wash* lights with a soft-edge beam of light and *hard* lights, which project a very defined beam of light and can project all types of patterns. These lights are primarily used for concert lighting and theatrical productions where movement of the lights and changing colors provide dramatic effects. These lights are used for face lighting only when they present the only practical alternative because of the useful remote focus capabilities.

Fluorescent Fixtures

Fluorescent fixtures have become quite popular in many studio installations. Utilizing the improved color temperature of compact fluorescent lamps, several manufacturers have developed a variety of useful fixtures. These systems offer some very compelling features such as low energy consumption, with minimal heat production. They are excellent soft or fill light but are somewhat limited in throw distances. Fluorescents are most commonly employed in news and other continuously on air forms of production to take advantage of their energy efficiency. Fluorescents are an excellent tool for the lighting director and can be effectively used in conjunction with standard incandescent fixtures.

Accessories

There are a number of accessories, which make positioning light fixtures more versatile. Despite the expenses and care expended on planning the grid system, it cannot satisfy every fixture mounting requirement; every studio must have some fixtures on *rolling floor stands*. In addition, many types of *clamps* and general grip equipment are available for special mounting situations.

One of the key accessories in the lighting system are *extension rods* either *straight* or *telescopic*. This device allows fixtures to be hung at a lower level than on the fixed grid. For example, a key light at a 30° angle (elevation) above the talent could be moved much closer to the talent and provide more light if it were hung lower. Counter-balanced devices such as pantographs and spring load telescoping hangers are not recommended because they cannot locked in place.

The *Century stand* has endless applications around the studio. It can grasp all types of material such as flags, cutters, cookies and support reflectors (see Figure 5.16-9). *Century* stand use is limited only by the lighting designer's imagination. No matter how many are available, they will all be put to good use.

Ultimately, the proper quantity of fixtures, accessories and other components of a lighting system for a studio of a given size can vary widely according to the requirements of a specific situation. For example, a studio with a fixed grid and a tight production schedule will function more efficiently with a heavy saturation of fixtures. With a large quantity of fixtures, major relocation of all the units is somewhat minimized. When sophisticated motorized grid systems are available, the ease of relocating fixtures will reduce the total quantity of fixtures for the same size studio. The choice of either of these two approaches is also dramatically affected by labor costs within a particular facility. Too often the expense of a few extra fixtures is not significant compared to the additional time and labor required to fully utilize a minimal complement of equipment or the cost of elaborate rigging. Final discussions on equipment purchase require a careful analysis of a station's specific production requirements in coordination with all the financial ramifications of various approaches.

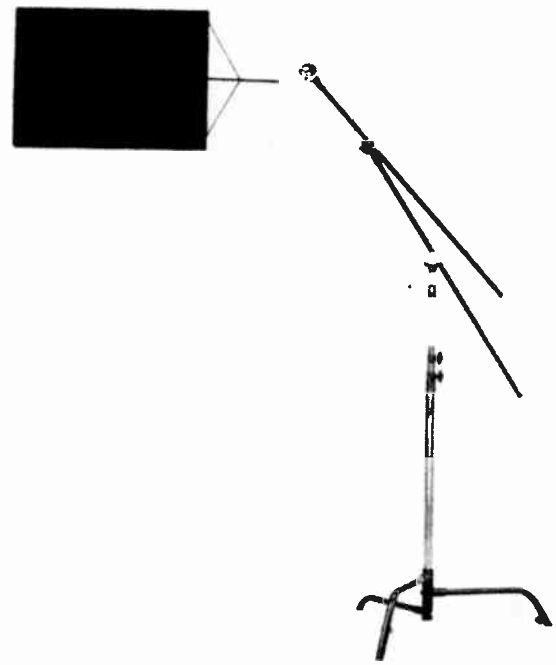


Figure 5.16-9. The three level legs allow the "C" stand to fold flat for storage. The head can grasp numerous flags, cutters, cookies, etc.

It should be evident from the previous general discussion of lighting problems that the close cooperation between the lighting designer and the video shader is absolutely essential. The video shader is in control of a marvelously versatile picture machine whose limits are stretchable in the name of art. Great television pictures are the result of careful teamwork by the lighting designer and the video shader.

STUDIO DESIGN

In designing a television studio, the lighting system is a major consideration as it is so closely interrelated to the physical size and shape of the room. While the television industry has seen tremendous technological improvements throughout its evolution, the basic phys-

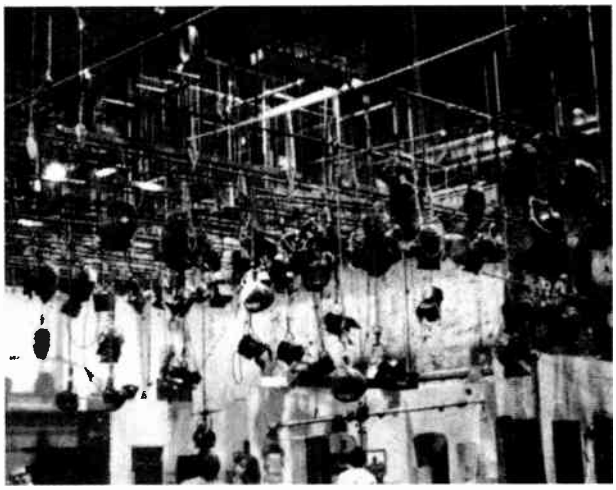


Figure 5.16-10. This large fixed grid studio illustrates the relationship of the various aspects of the studio lighting system. *(Courtesy of WPBT.)*

ics of light and its ability to describe three dimensional form cannot change. Consequently, many older studios have upgraded virtually every piece of broadcast equipment, while the basic components of an older, well designed lighting system still function extremely well.

Studio Size

Ideally, a studio provides the optimum environment for any type of production. In practice however, studios of various sizes tend to function best for particular types of production. The typical broadcast plant requires several sizes of studios to most effectively service its programming schedule. Typically, a small studio (1,200 sq. ft) is dedicated to news, interviews and public affairs; while 2,400 sq. ft is the most common size of small production studio. For general production, a studio of 5,000 sq. ft or more will offer fewer limitations. The size of the studio must be carefully determined by existing and future programming requirements, keeping in mind that the larger the studio, the fewer limitations. These are important decisions, since the lighting system must be planned relative to the size and specific requirements of the studio as shown in Figure 5.16-10.

Studio Height

When determining the height of a new studio, the lighting suspension system must be given careful consideration. The grid height for small and medium size studios is a function of the current fixed TV aspect ratio 4:3 and the normal wide-angle zoom lens. Most zoom lenses can cover approximately a +45° field. By calculating the actual width of this maximum horizontal dimension, the height of the grid can be determined by applying the aspect ratio (see Figure 5.16-11). Keep in mind that the studio lighting fixtures will hang approximately two feet below the suspension structure. In addition, a normal studio pedestal and cameraperson will prevent the lens from getting closer than approximately five ft in from a wall or any other obstruction, further reducing the maximum coverage of the lens. This method describes the theoretical maximum picture possible and the height the fixtures need to be mounted to stay out of the picture. However, television is primarily a close-up medium, and limitations in grid height can be overcome by various camera angles and special fixture mounting systems. High definition television will bring about a new, more rectangular aspect ratio (16:9) in the near future, which could actually reduce the theoretical studio height requirement.

Suspension Systems

The suspension system for the studio lighting is also a critical factor in determining studio height. All types of lighting suspension systems are commonly referred to as the *grid*. In its simplest form the system consists of a series of pipes suspended below the studio ceiling,

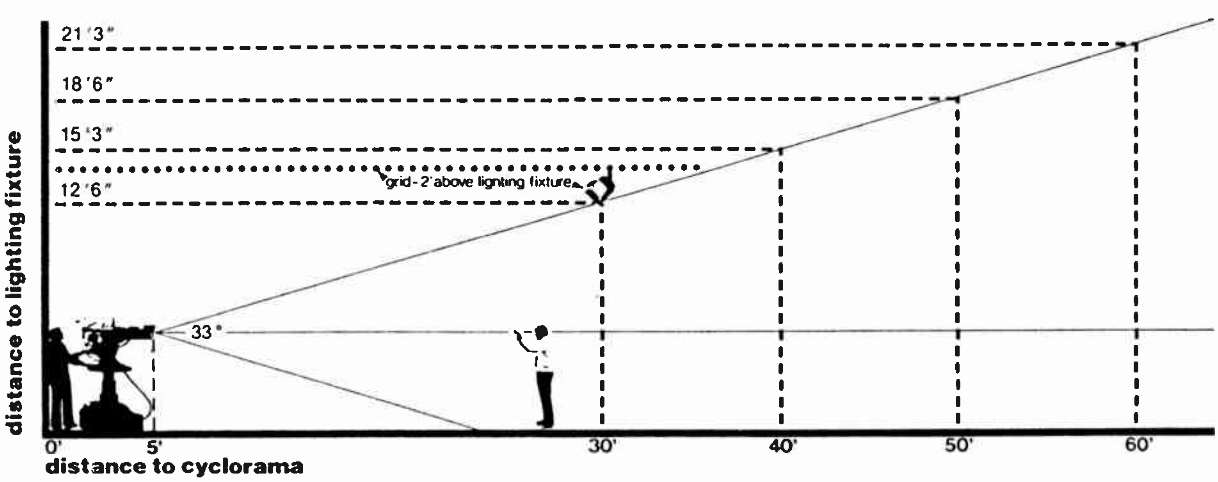


Figure 5.16-11. This drawing illustrates how the vertical aspect of a camera lens at eye level determines the height of a studio grid.

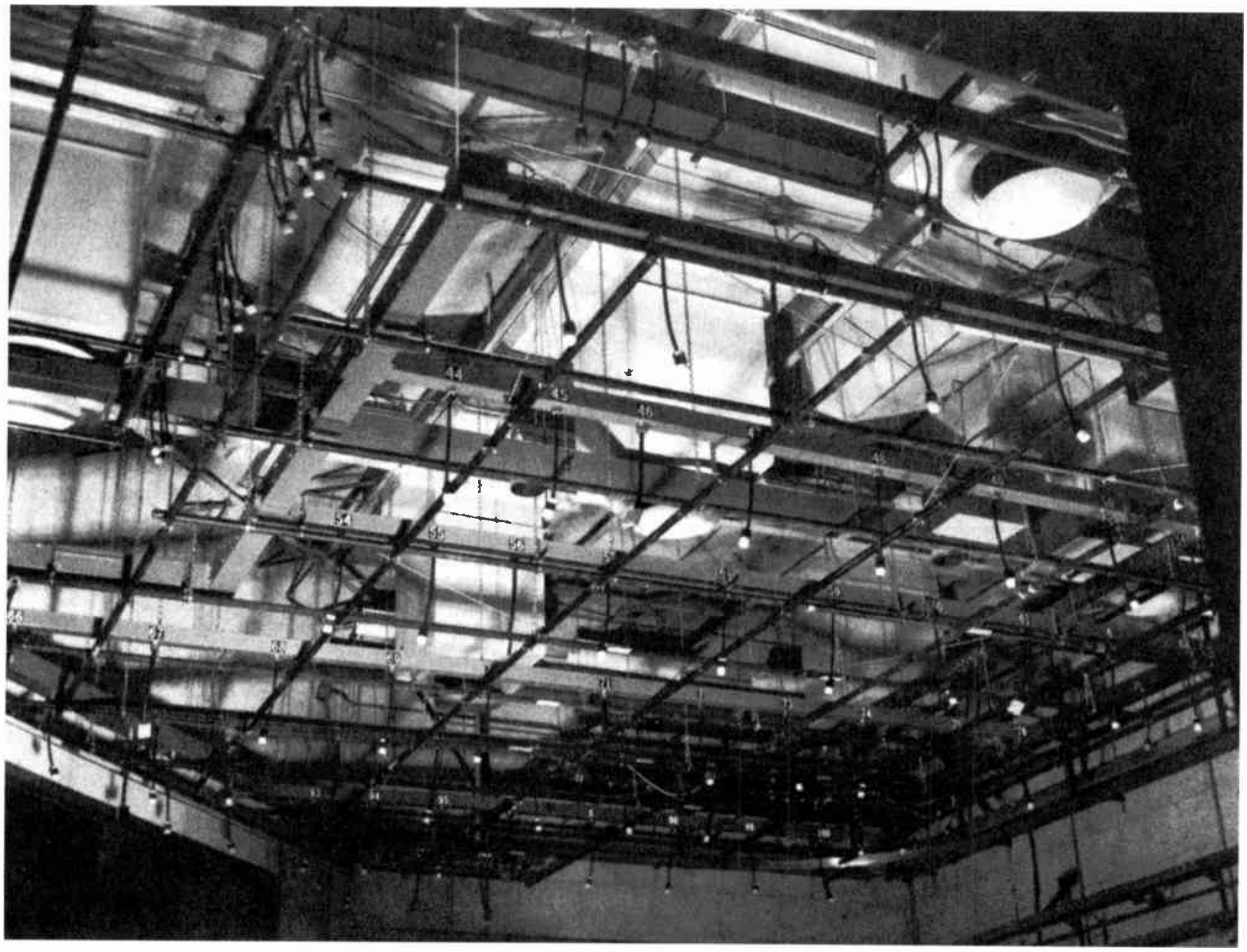


Figure 5.16-12. This fixed grid studio employs two levels to create a higher apparent background. Multiple pigtail outlets are positioned for cyc lighting units. *(Courtesy of WFAA-Dallas.)*

in a regular pattern. A grid must allow lighting fixtures to be hung anywhere over the entire studio. Actually the grid can consist of several different types or combination suspension system.

Fixed Grid

A *fixed grid* is the most common and least expensive system of suspending lighting fixtures (see Figure 5.16-12). The pipes are generally laid out on a perpendicular 4 ft × 4 ft spacing, which provides adequate flexibility in hanging positions. A common fixed grid height is 14 ft because it offers a good compromise between easy ladder reach and adequate clearance for wide shots. A fixed grid that is much higher than 14 ft increases the amount of labor needed to install and adjust fixtures. For a large multipurpose studio, a fixed grid could be a limitation, but the availability of a wide variety of self powered manlifts has allowed reasonable access to higher fixed grids. Bi-level fixed grids have been utilized to create a higher apparent background without additional cost and to keep the major portion of the grid at a reasonable height. To solve this problem more efficiently, several other systems have been developed.

Catwalks

Catwalks are generally also hung at a fixed level. Catwalks provide a measure of increased efficiency over a fixed grid but they are considerably more expensive. Catwalks allow studio electricians to work on lighting at the same time carpenters are handling the scenery. This can be an important time savings on a tight production schedule. Catwalks are usually arranged to create a fixed grid utilizing handrails in conjunction with an extension rod on each fixture. Catwalks when hung too low create the same height limitation as a fixed grid, but generally catwalks are employed at heights greater than would be reasonable for a fixed grid. Unfortunately, catwalks do not eliminate actually getting on a ladder to adjust the fine focus of each fixture. Catwalks primarily save time in setup and strike and offer the crews a safe work platform above the studio. In recent practice their high cost has prevented many installations.

Most studio fixture manufacturers offer developed pole-operated yokes for their units. Each fixture has a socket, which can be reached by a pole crank to adjust each of the basic functions, pan, tilt and spot-flood. There is a reasonable limit to how long a pole can be easily manipulated, but this feature considerably enhances fixed, high level grid systems. It is not possible to focus studio fixtures by this method as quickly or efficiently as by hand, but this system does provide a means of focusing otherwise unreachable units blocked by scenery and is seriously worth considering for any type of grid system.

Battens

The larger and therefore higher a studio that is required, the more complex the studio lighting suspension system becomes. In very large studios, a basic theatrical staging factor—"flyout clearance"—in addition to the TV aspect ratio, determines the grid height. Just as in a theatre stagehouse, flying scenery is a common requirement in larger scale TV production. Flown scenery commonly requires a minimum grid height of 40 ft. To raise and lower the scenic units and the lighting fixtures, a regular pattern of long battens or pipes run across the width of the studio. These battens run up and down on steel cables and are manually balanced by cast iron weights. These counterweighted battens solve the height problems but add additional work in rebalancing the counterweight arbors every time fixtures are added or removed from the batten. Counterweight rigging mounted along the studio wall encroaches on valuable floor space. Figure 5.16-13 compares the various approaches to the grid plan.

Winches

To solve the negative aspects of counterweight systems, many installations use electric winches to raise and lower the battens. These winches can be operated by sophisticated control systems which will allow the batten to be lowered for service or adjustment and then returned to an exact preset trim or level. The motors will easily lift any variable load within its designed capacity.

Many variations on the motorized batten system have been developed. In order to increase flexibility the battens have become shorter. This increases the number of motors required and the cost.

Modular Grids—Self Hoisting

To take advantage of the convenience and power of electric winches without excessive cost, a number of modular grid or short batten systems have been installed. These modules are essentially small pipe grid sections, which can be raised and lowered. Grids offer a flexibility similar to battens, but since they cover a larger area fewer motors are required. Both grids and modular short battens are designed to be self-hoisting. These systems greatly simplify structural requirements

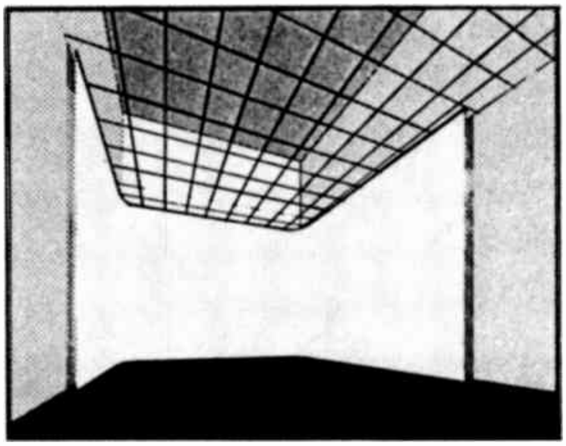


Figure 5.16-13(a). Fixed grid plan.

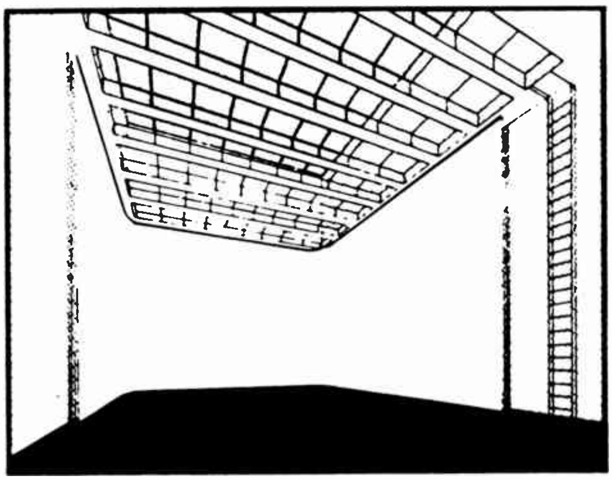


Figure 5.16-13(b). Catwalk grid plan.

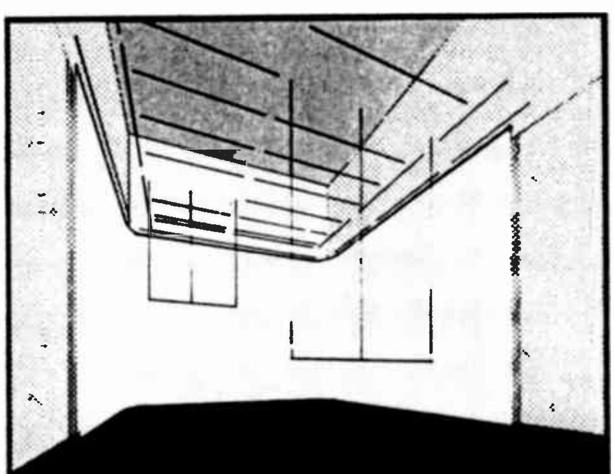


Figure 5.16-13(c). Batten plan.

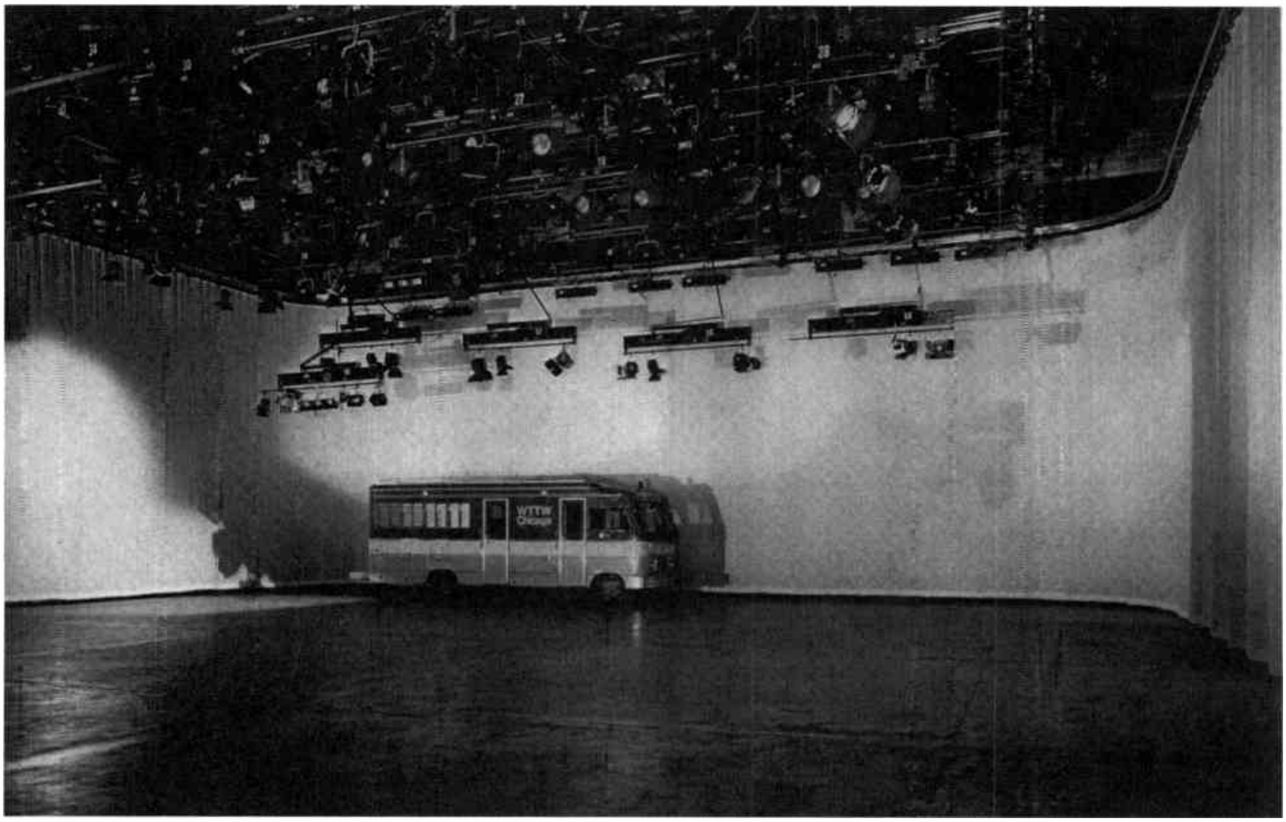


Figure 5.16-14. Self-hoisting batten are an effective suspension solution for large high grid studios *(Courtesy De Sisti Lighting.)*

of the studio, and the overhead structure remains clear of motors or steel hoisting cables (see Figures 5.16-14–5.16-18).

In studios which have either counterweighted or motorized battens a full walk-over grid should cover the entire studio. This grid-iron usually is made of steel grating or channels, as in a theatre stagehouse, and provides the support for the adjustable rigging. This walk-over grid provides easy access for any overhead suspension task any production may require.

In developing a fully integrated grid system, it is essential to coordinate all the building's structural, electrical and mechanical systems in relationship to each other and the grid. Whereas, in normal construction many of the mechanical elements in a ceiling are placed where convenient, it cannot be over-emphasized that improperly planned or installed ductwork and conduit runs can be a hindrance to production.

CYCLORAMA

Designing the cyclorama is integral in planning the studio lighting and grid systems. The grid or suspension system must provide a mounting position for the cyc lights located at the proper relationship to the cyc. Also, the total area of cyclorama will affect the calculation of the studio power service.

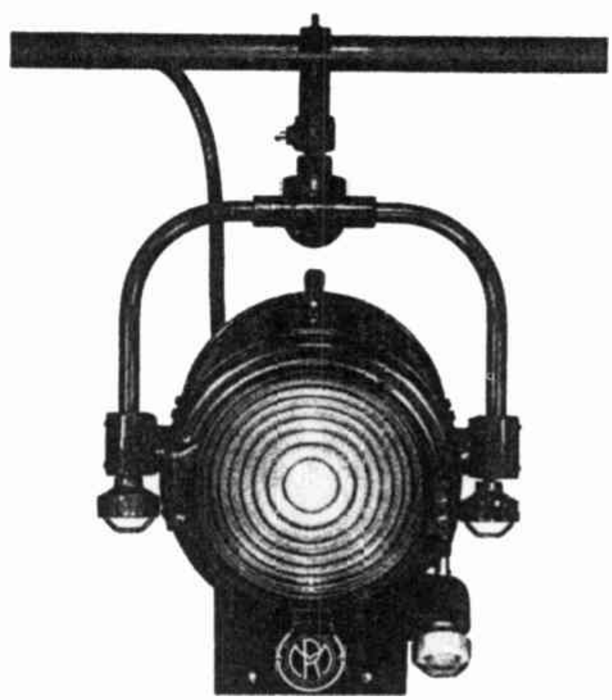


Figure 5.16-15. By inserting a long crank into the adjustment cups, this pole operated fixture can be panned, tilted or spot/flood focused from the floor. *(Courtesy of Mode-Richardson.)*

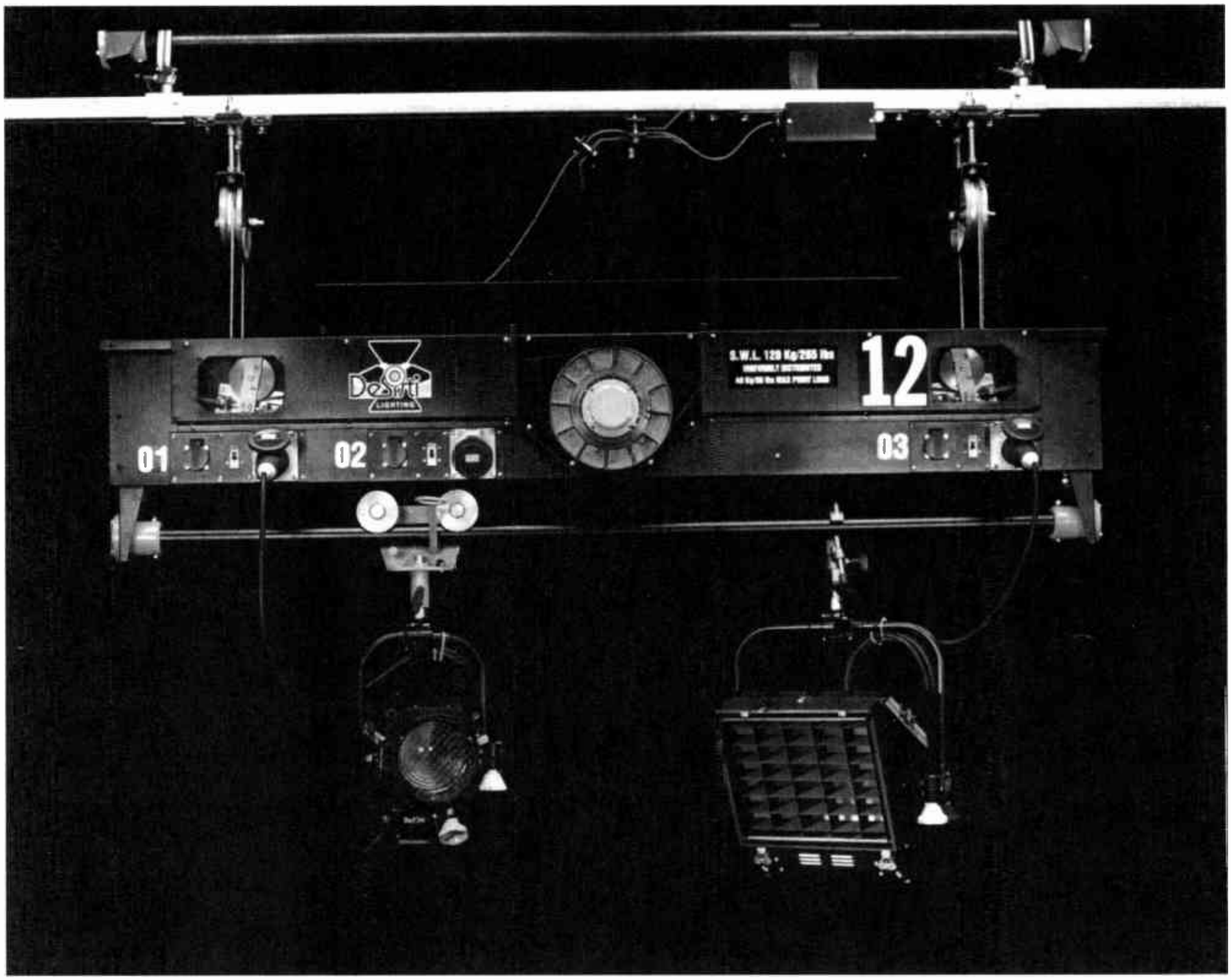


Figure 5.16-16. Manufactured self-hoisting battens are quick to install and efficient to use. *(Courtesy DeSisti Lighting.)*

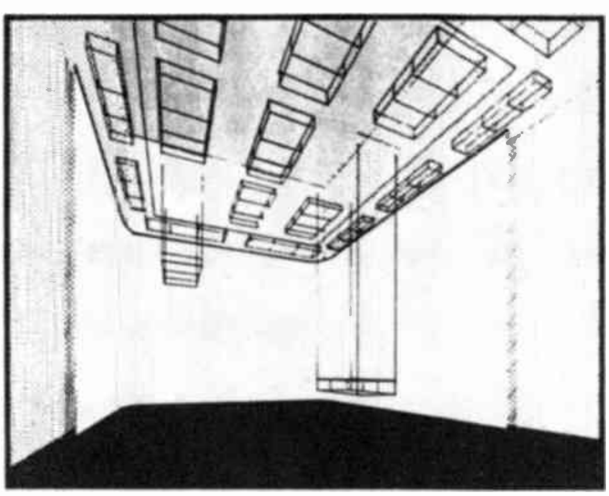


Figure 5.16-17. Modular grid plan.

Hard Cyc

Small studios often incorporate *hard* cycloramas which are smooth plastered surfaces that actually blend flush into the floor. They provide the ideal infinity effect which draperies cannot equally simulate. Hard cycloramas can be easily painted any color as needed and are especially effective for certain chroma-key techniques that can be spoiled by even the most invisible seams. Hard cycloramas are generally limited to small studios since their hard surface area invariably creates acoustical problems.

Cyc Pit

In very large studios a *cyc pit* is a compromise solution for creating the infinity effect (see Figure 5.16-19). The pit contains and conceals the bottom cyclorama lighting, which is essential for a tall cyc and the bottom of the drapery. When it is shot from the proper angle the pit will simulate a background without a horizon.

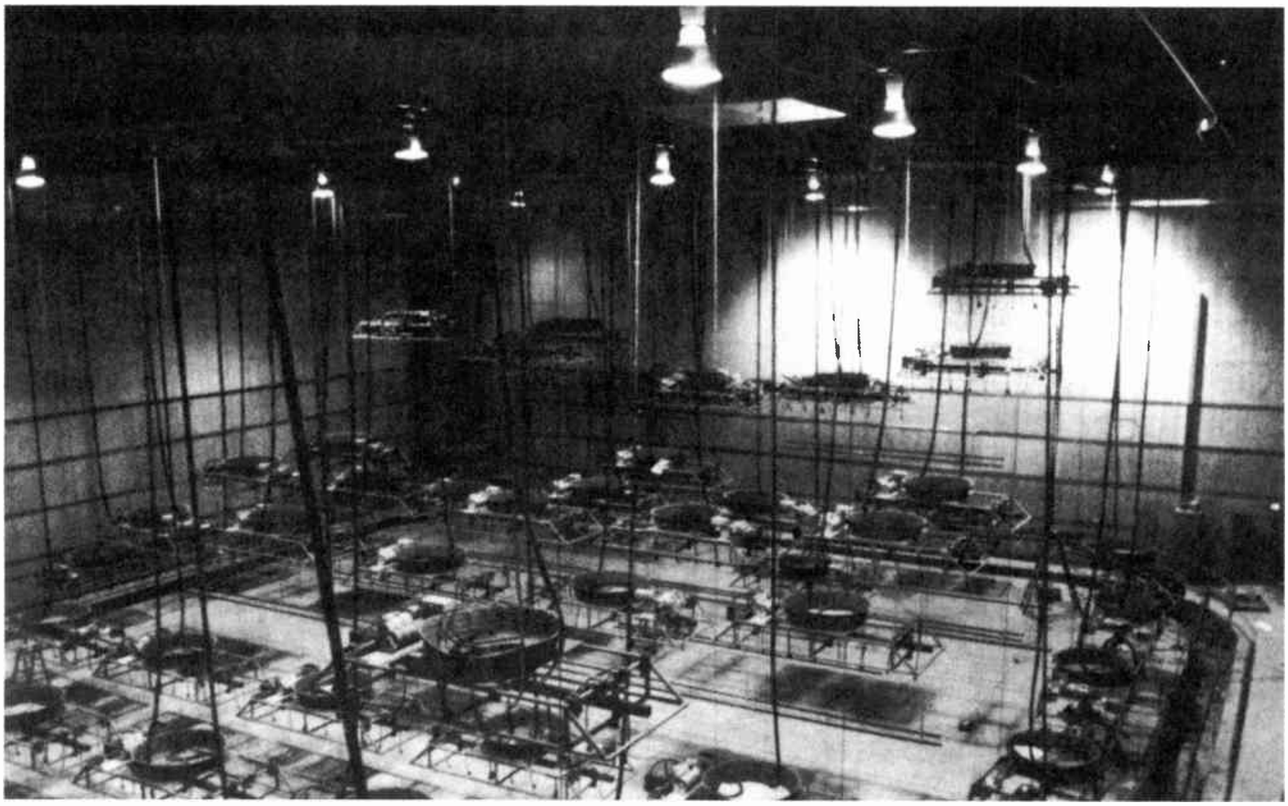


Figure 5.16-18. Modular self-hoisting grids allow variable height suspension of fixtures in a regular pattern through the studio. The round tubs serve as cable collectors for lighting and winch power. *(Courtesy Dallas Communications Complex.)*

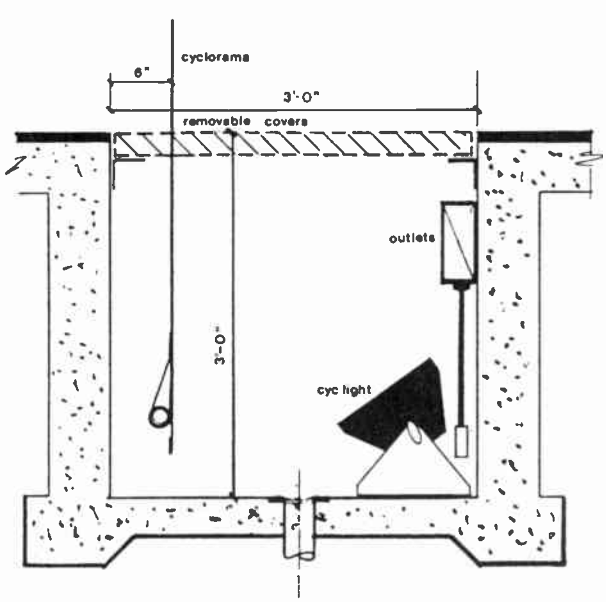


Figure 5.16-19. Cyc Pit: Section view through a typical cyclorama pit.

For most studios, a *drapery cyclorama* is the most convenient solution. This seamless drapery is hung on carriers which roll along a track and allow the cyc to be positioned anywhere around the perimeter of the studio (see Figure 5.16-20). Two parallel tracks permit another type of background to be pulled in front of the stretched cyc. Switches on the track system allow the draperies to be easily transferred to the front or rear tracks. *Leno* or *filled scrim* is the most common drapery material, although seamless muslin is an inexpensive alternate material. Translucent plastic rear-projection screen type material is also a popular material because it can be lit from either the front or the rear.

In large studios where experienced lighting directors are available, true white cycloramas are used to achieve greater color intensity on the cyc. However, in smaller studios, where the talent occasionally must work close to the background or where limited control equipment is available a 60% TV white cyc should be used for better control of contrast.

Drapery cycloramas are generally furnished with jack chain weights. Removable pipe weights bent to match the shape of the track should also be provided to create a wrinkle-free background. One of the most common errors in cyclorama design is an insufficiently large radius at the corners of cyc. No matter what material the cyc is made of, the larger the radius the easier it is to light evenly and accomplish the desired effect.

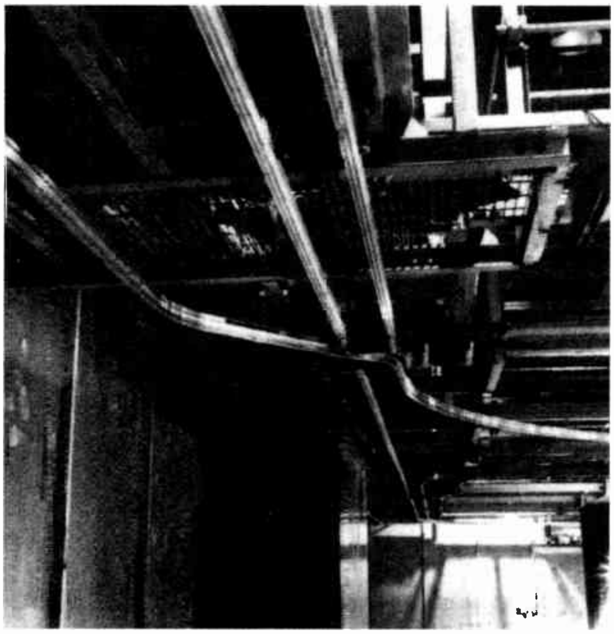


Figure 5.16-20. Cycloramas and draperies can be shifted to various track configurations by utilizing transfer switches. (CBS-NY) *(Courtesy of Peter Albrecht Corporation.)*

Cyclorama Lighting

Generally the arrangement of doors into the studio will define the most functional area for the cyc to be normally positioned. In the basic design of the grid or suspension system the type of cyc lighting system should be predetermined, and the proper hanging system provided accordingly.

There are several types of cyclorama lighting fixtures—striplights and a type of widespread fixture commonly know as *Far Cycs* (see Figure 5.16-21). Striplights are continuous rows of quartz halogen, MR-16, or Par lamps, which for an 18 ft high cyclorama should be mounted 5 to 6 ft from the cyc, while

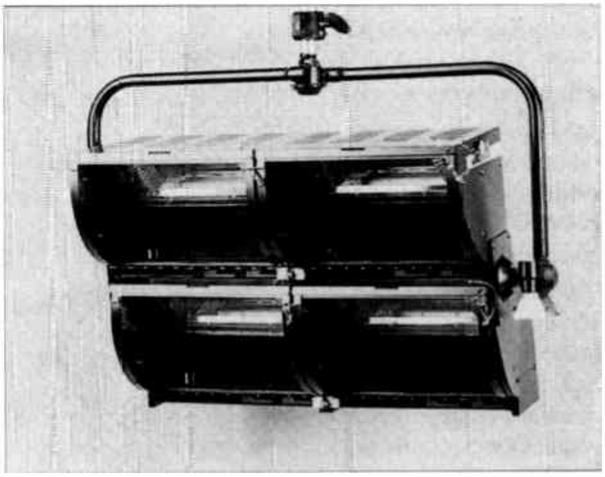


Figure 5.16-21. Four unit Far Cyc fixtures are very effective in lighting tall cycloramas. *(Courtesy Desisti.)*

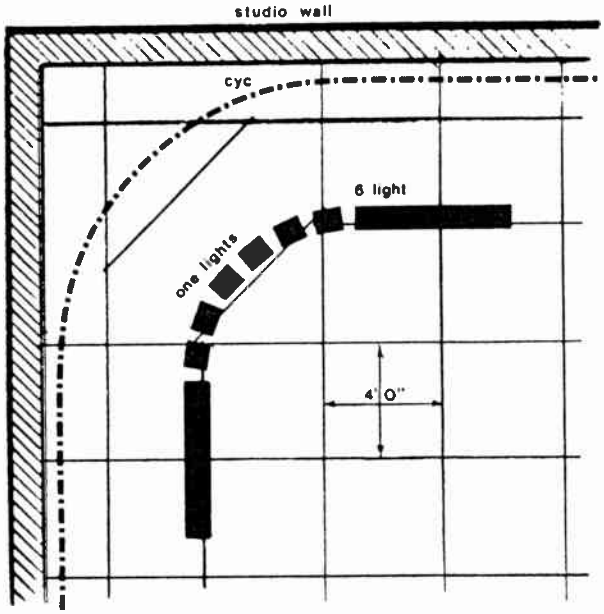


Figure 5.16-22. Cyc Light Layout: Typical arrangement of strip cyc lights relative to the cyc curve.

the far cyc units should be mounted 7 to 8 ft from the cyc (see Figure 5.16-22). Another type of fixture is referred to by its brand name *Pallas*. It is a version of a striplight that is made up of individual cells which can be placed to match any curve or straight surface. The entire suspension system should be designed around these dimensions. Far cycs generally will light the cyc as evenly as striplights, with less wattage. Because far cycs are mounted a greater distance from the cyc, they force the talent further away from the background. Generally, striplights should be limited to three colors, otherwise the separation between alternately colored lamps is too great to provide even coverage.

ARCHITECTURAL CRITERIA

Electrical

With the basic studio size determined and the net production area (NPA) defined by the cyclorama, it is possible to determine the power requirements for any given studio. The power requirements for studio lighting are a direct function of area and the required level of illumination. This power requirement remains consistent regardless of the grid height. For a lower grid, a greater quantity of smaller wattage fixtures are used, while a higher grid will require fewer fixtures of increased wattage. In either case, the total watts per square foot will remain roughly the same.

An average of 55 W per square foot of NPA has been proven in production to provide sufficient power for any normal television lighting requirement. This method of calculating the studio load will provide sufficient power for virtually any situation. This is generally more power than is actually required for the average studio with today's state-of-the-art equipment, but it allows for over-lighting by novices and higher levels necessary for special situations. It also provides sufficient power for an average cyclorama as well as lighting the entire studio as required for audience participation. Very tall cycloramas which may required double rows of fixtures at the top, bottom or both, cyc lights should be calculated as an additional power requirement based on the wattage per lineal foot of cyclorama according to the lighting system chosen.

This calculated power service describes a real maximum probable load, and the feeders must be able to supply this full amount of power. Only certain limited productions will ever require this full amount. Also, note that for dimmer-per-circuit systems, the dimming capacity will be far greater than this calculated power service. The full dimming capacity need not be fully serviced, since the larger number of dimmers is a matter of convenience, and should never be fully loaded beyond the maximum probable load.

Air Conditioning (HVAC)

Because the maximum lighting load seldom occurs, it is unnecessary to use that maximum load as the basis for the air conditioning capacity. Production practice has shown that a diversity of 60% can be applied to the maximum load and still provide sufficient capacity for full period shooting. Of course any other heat generating devices and the population of the studio should be included in the air conditioning calculations. A properly designed HVAC system will require very large ducts to meet stringent acoustical requirements. These large ducts can often interfere with a grid system and should be closely coordinated.

Determining Studio Power and HVAC Requirements

1. Net Production Area (NPA) =
   NPA = Actual studio area in sq. ft minus area behind cyc and other areas unusable for production.
2. Approximate Number of Studio Lighting Outlets Required:
   $$\frac{\text{NPA}}{\text{16–18 sq. ft}} = \text{Number of outlets required (dimmer-per-circuit).}$$
   NOTE: Actual number of outlets determined by specific layout.
   NOTE: Breakdown of 20 A circuits and 50 A circuits varies by grid height.
3. Studio Lighting Load:
   NPA sq. ft × 55 W = Total W
4. Studio Lighting Power Service:
   $$\frac{\text{Total Watts}}{\text{120 V}} = \text{Total W}$$
   $$\frac{\text{Total Amps}}{3} = \text{Power Service 3 phase/4 wire 120/208 V}$$
   NOTE: Round off to the next larger standard panel size.
5. Studio Lighting Heat Load for HVAC:
   Studio lighting load (kW) x 60% (Diversity) = Heat Load for HVAC
   Design (kW)
6. Dimmer Room Heat Load for HVAC:
   HVAC Heat Load (Item 5 above) × 5% = Dimmer Room Heat Load

Electrical Distribution

For a studio to be totally flexible, lighting equipment power must be distributed uniformly throughout the studio. At the grid level, power is commonly distributed through prewired plugging strips. These strips are mounted directly to the grid, catwalks or fly in and out on battens or moveable grid sections (see Figure 5.16-23).

Each circuit terminates in the studio in a pigtail/outlet. There are two types of connectors in common use—*stage pin connectors* and *twist lock*. Stage pin connectors are less expensive and more common in rental equipment. If you envision renting additional fixtures on occasion, this may be an important consideration. In addition, the cost savings of stage pin connectors recur with each fixture and cable purchase. Twist lock connectors have a positive locking feature. The final choice should be based on your specific requirements.

The number of circuits and their capacities is also related to studio size. The actual number of circuits of based initially on the square footage (approx. one outlet every 18 sq. ft NPA) and then altered as necessary to conform to the particular grid system and the cyclorama layout. Dedicated circuits for the cyclorama are often overlooked. The cyc lights consume a large number of circuits and once they are hung in place they will seldom, if ever, be moved. Because of the rather wide spacing of the cyc units it is sometimes more efficient to feed these lights from individual grid

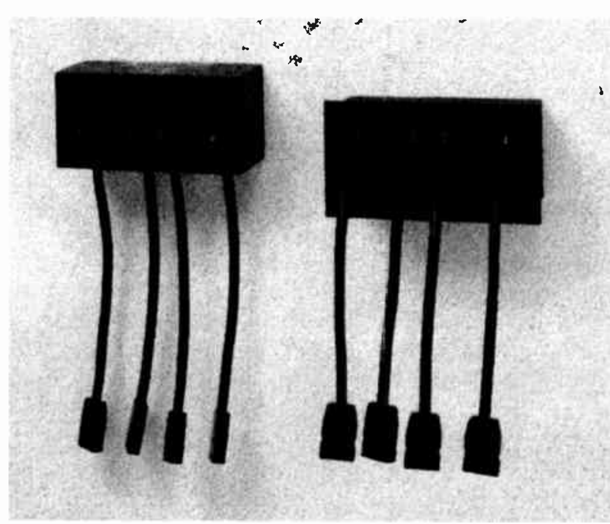


Figure 5.16-23. Wall mounted outlet boxes are generally located around the perimeter of the studio. *(Courtesy of ETC.)*

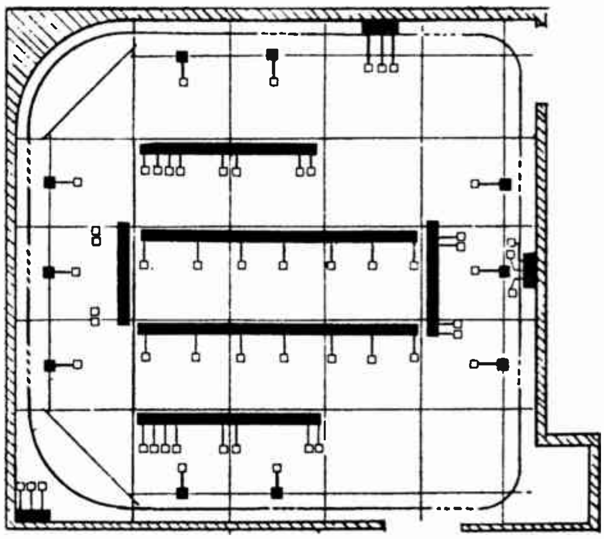


Figure 5.16-24. Typical studio distribution plan for a small industrial studio.

mounted junction boxes rather than a plugging strip. Also, when using far cyc lights it is often convenient to double up the outlets to take full advantage of the 20 A capacity of the dimmers.

Additional circuits should be located around the perimeter of the studio at grid level (see Figure 5.16-24). This is a natural backlight position for a set facing away from the wall. Properly located circuits save considerable time in running jumper cables.

At 30 in. above floor level, around the perimeter of the studio wall, mounted outlet boxes should be provided. Generally, the governing factor for their placement is relative to a layout of floor mounted cyc strips. Otherwise these outlets are used for miscellaneous lights on floor stands and practical fixtures on the set.

For most studios the vast majority of circuits will be rated 20 A. The larger the studio, the greater the quantity of 50 A circuits. In a medium size studio (3,500–5,000 sq. ft), 50 A circuits are generally located in a regular pattern throughout the center area of the grid and slightly more frequently around the perimeter backlight position. In larger and higher studios the density of 50 A circuits must be increased, although the overall outlet density should not be significantly decreased. On adjustable height grids, a full complement of both 20 A and 50 A outlets should be provided for using the appropriate fixtures for the varying heights. Large studios (10,000+ sq. ft) with high grids must be furnished with 100 A circuits.

Lighting Control

The importance of a dimming and control system cannot be underestimated. Dimmers allow easy control of numerous fixtures, balancing and recording of levels and the blending of colors. A dimmer system frees the lighting director of the unnecessary burden of calculating and controlling the loads through more labor intensive mechanical methods. Electronic dimming and control allows the execution of complex lighting cues, which are a very effective production element.

Most studios today are outfitted on a basis of one dimmer-per-circuit. In this type of system every circuit terminates in its own dimmer, with the integral circuit breaker protecting both the dimmer and the circuit. The dimmer-per-circuit system gives the lighting director individual control of each individual fixture, or group of fixtures plugged into that circuit. Normally, the studio is outfitted with 20 A and 50 A dimmers.

Dimmer Bank

The individual dimmer units generally plug into electronic equipment racks to form the dimmer bank (see Figure 5.16-25). This modular system also allows quick plug-in substitution of faulty dimmer modules. Depending on the manufacturer, up to 96 individual 2.4 kW dimmer modules can fit into a single rack. Most modern studio quality electronic dimmers utilize silicon controlled rectifiers (SCR) although a few have used MOSFET (see Figure 5.16-26 and 5.16-27). These units are very reliable and are universally available in 2.4 kW, 6.0 kW, and 12.0 kW (20 A, 50 A, 100 A @ 120 V) ratings. Only dimmers which have sufficient filtering to prevent unwanted RF interference and excessive filament vibration should be considered for use in television studios. Toroidal chokes are uti-

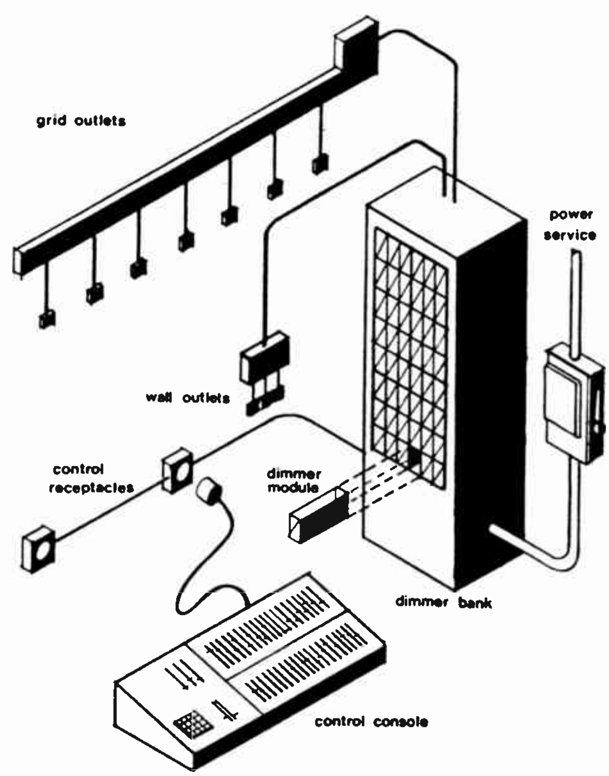


Figure 5.16-25. Dimmer system riser diagram; indicating the basic components of the dimmer system.

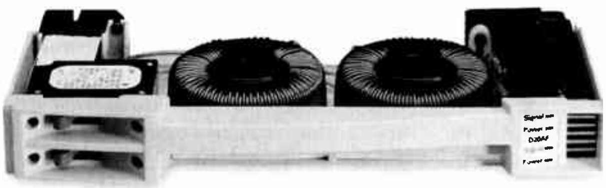


Figure 5.16-26. Tray mounted dual dimmer modules allow high density and quick replacement. *(Courtesy of ETC.)*

lized to control the rise time of the dimmer and the associated electrical noise.

The location of the dimmer bank is an essential part of the initial studio space planning. The dimmer room should be centrally located to minimize the length of all the circuit homeruns to avoid voltage drop and excessive installation cost. This room should be sized to allow sufficient space for required conduit radii, access to the feeder lugs and adequate front clearance as specified by the local code.

Virtually all SCR dimmers are approximately 3 to 5% inefficient. Therefore, they could create heat equal to 5% of the energized lighting load. Since the maximum lighting load is an infrequent occurrence the dimmer room cooling should more reasonably be based on 5% of the diversified load on which the studio HVAC is based.

Control Consoles

In electronic dimming systems, each dimmer circuit represents a channel. Each dimmer channel can be programmed to retain a particular dimmer level setting. A group of preset levels is called a *scene*. While manual control panels with two scenes are sufficient for small productions, even a relatively simple production could require many more. The physical size of manual multiscene control panels is cumbersome.

Even the most modest control consoles offer enough control channel for individual settings of every dimmer although it is possible to electronically patch dimmer circuits to smaller number of control channels. Although the simplest systems physically resemble a standard manual two scene preset system, they have virtually all the capabilities of the largest systems systems. Each slider can represent any channel, group of channels or act as a submaster, rather than a hardwired individual dimmer. The specific number of control channels is a matter of the physical electronic design of the console. Obviously, more channels afford greater control within a single scene.

In a dimmer-per-circuit system it is not unusual for a medium size studio to have in excess of 200 dimmers. To control this large quantity of channels it is possible to patch the dimmers into to control channels. The patching function allows any dimmer to be controlled by any channel. For example the blue cyc lights load may require 12 separate dimmers. By patching them into the same channel they will operate together in perfect unison. All this patching occurs within the console and does not require any cord, plugs, or diode pins. For small to medium size studios, consoles with over 100 miniature potentiometers in 2 or 3 presets are available to provide the operational convenience and simplicity of manual systems.

After patching the dimmers to channels and setting the desired levels, the entire preset can be stored in a memory by assigning a memory number. This preset can then be recalled by keying the appropriate number or operating the slider to which the preset has been assigned. This type of system is very economical and is suitable for most small and medium studio situations.

For studios which encounter more complex production requirements there are a number of systems which resemble a personal computer with specialized keypad and controls with much greater capabilities (see Figures 5.16-28 and 5.16-29). One type of system which is common to several manufacturers utilizes a single or double color CRT or flat panel to display various functions. In operation, the screen displays the channel numbers and below each channel is a two digit number for the intensity. At the bottom of this field of numbers is the *cue sheet* which displays various operational functions such as cue numbers, fade times, etc. The most complex systems employ dual displays to actively display more information. These systems offer many features and can even tell whether the lamp is burned out from a console readout.

The control panel on this type of computer control generally has a series of submaster sliders, X, Y faders, numerical keypad and a function keypad and some tactile type of encoder for easily altering fade times and intensities.

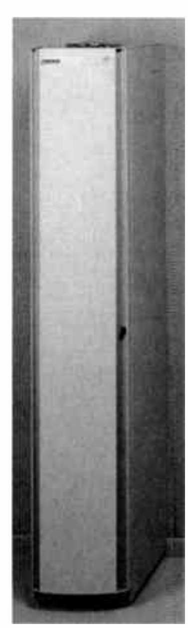


Figure 5.16-27. Dimmer systems are packaged in extremely compact rack configurations for every size studio.

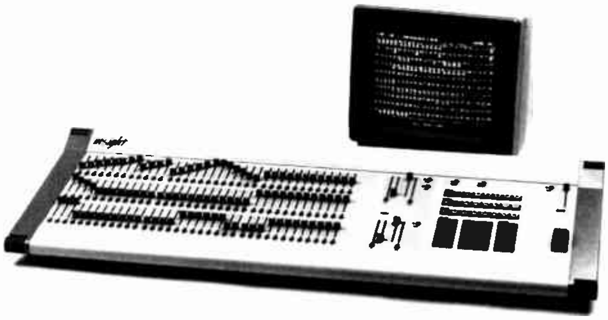


Figure 5.16-28. For medium size studios, memory consoles offer both manual and memory control. *(Courtesy of ETC.)*

Computerized lighting systems store their memory on one of several types of removable storage media for reuse. This permits a complete copy of all the settings in memory. In addition most larger systems are available with varying degrees of internal backup memory systems and can be networked to interconnect various types of remote controls and peripherals. The operational software is permanently stored in ROM within the machine.

Most computer lighting control systems were originally designed for the legitimate theater where the easy daily repetition of very complex multipart cues is the primary goal. Some manufacturers have modified their programming to be more sympathetic to television's somewhat less predictable demands. In selecting a system with this caliber of sophistication, the lighting director must carefully determine which features are really necessary.

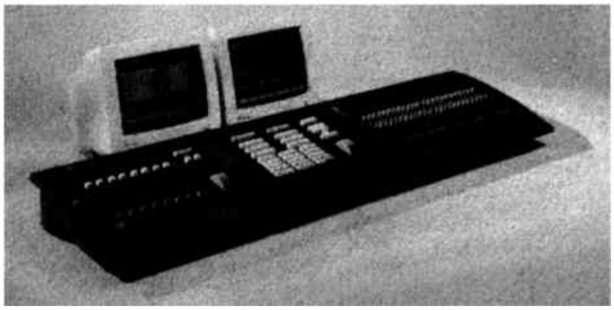


Figure 5.16-29. The most sophisticated consoles utilize dual CRT displays and can control very large quantities of dimmers with equal number of discrete channels.

Computerized lighting systems have greatly simplified the installation of the control wiring. While analog systems required at least one wire for every dimmer, these computerized systems use standard computer network wiring. An advantage to this simplified cabling is that the console can be easily relocated to any of the plug-in stations. In addition, when the main computer is located outside the studio, a small focus remote the size of a handheld calculator can be used in the studio to activate fixtures as necessary for focusing or other simple operations. The technology of computer networking is now used on lighting control systems to add greater flexibility and multiple control locations. This allows the console to be located in the most convenient position for a particular phase of the production. It also allows several studios to share a more sophisticated system as required, since the largest systems can plug into the same control wiring as the smallest.

Computerized control has made sophisticated remotely controlled fixtures possible. Most consoles now have the ability to control moving lights as well as the dimmer system from a single console. A single console can be used to program and record the intensity, position, speed of movement, iris, internal pattern selection, color mixing, pattern rotation as well as soft or hard beam patterns. The console records and replays all these functions as desired by each cue programmed.

THE PLANNING PROCESS

Most television production professionals have the opportunity to be involved in planning a new studio maybe once or twice in their entire careers. Even years of experience in studio production are not necessarily the best preparation for coordinating studio requirements into a construction process. Often, a new studio provides an opportunity to acquire a complement of equipment, which is more sophisticated than the existing staff's level of experience. In this situation, an experienced lighting designer should be consulted in preparation of the equipment requirements and evaluating the quantities required. While manufacturers are sometimes helpful in this area, they are still primarily interested in selling their product, and no single manufacturer offers a full line of suitable equipment in every area. By working with a consultant, you can share their experience in the planning and design of the lighting system for your studio.

5.17
REMOTE VIDEO PRODUCTION

NED SOSEMAN
KCTV, MEREDITH BROADCAST GROUP AND
CARL BENTZ
INTERTEC PUBLISHING, OVERLAND PARK, KS

INTRODUCTION

Television brought the dimension of pictures by radio into our lives. Yet, over the years, technology has achieved technical successes almost as important as was the mere existence of early telecasts. Once bound to a studio by sheer size of equipment, TV production is now almost at home in the field. Going remote is more than just buying battery-powered equipment, however. This overview of field production will examine some of the aspects that impact the development of outside broadcast capabilities.

WHAT IS FIELD PRODUCTION?

Producing television at remote locations is like piloting a jet fighter aircraft. Both are filled with elements of the unknown, danger and excitement, and both conclude with a distinct sense of accomplishment or disaster. Both depend on sophisticated skills, equipment and support and both must prepare for the eventuality that anything that can go wrong probably will.

The secret to technically successful remote video production is to rehearse and prepare for the worst. Like fighter pilots, experienced producers have learned that practice and standardized procedures are the keys to consistently successful completion of the mission at hand.[1]

Safety First

While Chapter 9.5: *Electrical Shock* reviews electrical safety issues thoroughly, no discussion of remote video production would be complete without emphasizing the fact that electricity kills. Too many people have been killed by electricity during remote video productions.[2]

The prevailing cause of remote production deaths is overhead power lines. An ENG vehicle's 40 to 50 ft pneumatic microwave antenna mast can easily contact overhead power lines. The first rule in remote video production is "look up before raising the mast!" and look around for other potential electrical hazards, including standing water, before doing anything else. No deadline is worth death.

Follow local electrical codes. Regularly inspect all electrical connectors and cables for physical damage and loose connections. Keep splice type connections dry by wrapping connectors as well as a few inches of cable on each end with plastic film then sealing the film with gaffer's tape. Electrical problems can be catastrophic, and there is always the chance that a fire marshal might want to inspect the location.

A host of other dangers, from falling objects to fires, also lurks in the field. A camera might crash to the ground if the tripod legs are not properly set. A forklift might run over a cable and short or break it. Wind might blow a hot unsecured light tripod into dry brush and start a fire. If it could possibly happen, change or fix it. Avoiding potential danger (and legal difficulties) is the top priority for everyone involved in remote production.

Production Categories and Styles

Video production work performed beyond the confines of four studio walls is considered remote or field video production. The emergence of portable video cameras and VTRs has produced two categories of remote video production: electronic field production (EFP) and electronic news gathering (ENG).

As the ENG name implies, this is the *shoot it as it happens category*. Typically, crews are minimal, retakes are few and time constraints are tight. It is usually the default category for fast or late breaking daily news stories.

The ENG category is defined by the lack of control a production crew has over the subject in the frame. As a result, ENG equipment must be adaptable to function in a variety of climatic, shooting and lighting situations, and able to do so with battery power. Typically, an ENG production system is operated by a one person crew.

Remote production outside the fast news lane is known as EFP. This category is characterized by control. EFP shoots are better planned and as near studio-quality as possible. For instance, major outside broadcast tractor trailer rigs with multiple cameras and sophisticated live audio and video switching facilities would define one end of the EFP spectrum. At the other end would be a high quality single camera, a lav mic and an artificially well lit scene. In general, the

---

[1] Taibbi, Steven, "Shoot the Video!" *Videography Magazine*, Miller Freeman P.S.N. Inc, New York, Volume 20, Number 2, pp 32, 34, February 1995.

[2] Bell, Mark, "Truck Safety Remains an Issue" *TV Broadcast Magazine*, Miller Freeman P.S.N. Inc, New York, pp 52, 54. March 1996.

difference between ENG and EFP is time, money and control.

The ENG category has spawned its own shooting style, characterized by a live or reality feel, which, of course may not be either. This shooting style is achieved with a single camera, hand held or mechanically supported and minimal or harsh lighting augmentation. While it is possible to shoot ENG style with multiple cameras, or even ENG style during an EFP shoot, the ENG category should not be confused with ENG style.

Whether using the most sophisticated multi-camera production trailer, an ENG van or a camcorder in the back seat of your personal auto, the discipline of video production remains the same. Only the scale changes.

PLANNING THE SHOOT

Many well worn stories have been spun describing scenarios where a project with more than one leader resulted in problems. Similarly, successful video productions (in the field or studio) can have only one person in charge. All members of the crew must understand and follow the direction of the person in charge. Typically, that person is either the producer or director.

The producer is responsible for the entire production. Typically, producers visualize, secure financing and organize the production. The producer approves the scripts and helps create the storyboards. Some producers like to direct, most use a contract or use a staff director to execute the plan. At the point where a director is brought into a production project, the director becomes the boss—in charge of fulfilling the vision of the producer. Electronic equipment and production crew responsibilities are usually delegated to a technical director (TD) or engineer-in-charge (EIC).

Before the first frame of video is shot, the producer must define the scope of the project on paper. Typically, the first documents created are the budget and the production schedule.[3]

Budgets, financial and time, govern the options of any producer. Outside of the studio walls, everything becomes more difficult and expensive to control. Items taken for granted in a studio environment, such as parking, restroom facilities and telephone communication must be identified and included in the production plan.

Does the production call for helicopter shots of thousands of extras on horseback galloping across the desert, for intimate ENG style reality shots inside a moving Miata, or both? Does the production call for overseas travel or is the shoot across the street? Regardless of the size of the production, the planning process remains the same. The size and budget controls the scope.

If production crews work for more than a few hours, the cast and crew get hungry. To avoid wasting time, many producers opt to cater food to the location. Good catered food raises morale and keeps the cast and crew on location. Ample drinking water and cups are also necessary to keep crews focused.

The producer and director must also develop contingency plans for potential problems, such as bad weather and equipment failures, before the production begins.

Once these questions have been answered, and talent has been secured, the producer generates the working production schedule. The schedule is a timeline that lists the stages and steps of the production process from the first day of preparation until the project is completed in post.

Choosing Locations

Using the production schedule, the producer can identify the crew and logistic requirements. A one person crew may be adequate to shoot some field video or pick up sound effects. Other scenes may require grips, boom mic operators, audio recordists, camera operators, cable pullers, focus pullers and others. The budget usually dictates crew size.[4]

If the production venue is out of town, the producer is also responsible for securing travel arrangements and accommodations for the cast and crew. Of course, before any arrangements are made, the producer must scout the location. Scouting includes shopping for crew accommodations as well as performing a thorough aural and visual check of the actual shooting locations.

Every field shooting location is unique. In addition, the technical aspects of virtually every shooting location change with the passage of time. Thus, the site that worked well last year may not work so well this year. Similarly, the site that seemed perfect when it was checked at 11 a.m. may have a different level of acceptability at evening rush hour. Check the site at a time similar to the hours the shoot will occur.

There is much more to a location than its visual attributes. The level of ambient noise may be too high or may change at different times during the day. There may also be sources of natural or man-made noise, such as a waterfall or portable power generator, that must be avoided. If noise cannot be avoided, then audio fix up work, such as sweetening and dialog replacement must be planned and budgeted for post.

Depending on the scope of the production, local shore power sources may or may not be a consideration. While using shore power may seem convenient, most standard wall outlet circuits will not support the power professional video production gear consumes. Often, in cases where quartz lighting is used, a portable generator may be the best source of adequate power. If shore power is the only choice, identify the location of the circuit breaker box, find the specific breaker on the circuit used and verify the current rating of that circuit breaker before doing anything else. A total

[3] Dellaria, David, "The Ten-Minute Miracle", *Videography Magazine*, Miller Freeman P.S.N. Inc, New York, Volume 21, Number 5, pp 68–70, 72, 74, May 1996.

[4] Ibid.

circuit load exceeding 65% of the circuit breaker's current rating is virtually guaranteed to cause the breaker to trip over time.

When possible, it is best to use separate circuits for electronics and lights. Some lights may have a tendency to generate noise spikes on the ac power line and some electronic equipment may tend to be susceptible to such spikes. Quartz lights, when powered over time, tend to heat and trip circuit breakers operating near their maximum capacity. Keeping the lights on a separate circuit will keep lighting and electronic problems separate.

If renting a generator is an option, rent a unit that can supply at least double the total load the production equipment will draw. Shop for a well maintained generator. Inquire about the condition of the generator and ask to see its maintenance log. Do not rely on a generator that is nearly due for scheduled maintenance or one that has no maintenance record. If the budget permits, rent a backup generator.

Bringing fixed frequency two-way radios and wireless mics to remote locations can also be treacherous. Unless you are operating in your home city, other people may be using your frequencies. Check with the local SBE frequency coordinator to ensure your production will not be plagued by interference and that you will not be interfering with anyone else. Frequency agile RF equipment helps avoid interference problems.

With the location identified and scouted, and production schedule in place, personnel requirements would be the next element to address. Working together, the technical director and producer or director must identify all positions to be filled and develop a crew assignment schedule. Typically, crews are scheduled to report to the location at least 30–60 minutes before the shoot begins, not including extra hours or days for setup, checkout and teardown.

Equipment

The next planning step is to develop a list of production equipment requirements. A physical checklist is the best method to identify every possible item that might be used in any shoot. Working backwards, the technical director can use the checklist to determine what gear is not necessary. What is left is what is needed.

Remote production equipment can be categorized in four equally significant groups: Video, audio, lighting and communication. (A fifth category, microwave, is usually reserved for ENG work.) Among the categories, communication has gained notoriety as the most important and most likely to fail. Unfortunately, because of its behind-the-scenes invisibility, it is also most likely to be ignored—until failure brings an entire production to a grinding halt.

Modern communication systems are the backbone of the production. The larger the crew or venue, the more critical communications become. Communications include telephones, two-way radios and crew intercom systems.

Before embarking on any remote production verify that all necessary communications equipment, wires and headsets are on hand and in good working order. Further details on communication systems are covered later in this chapter.

Typically, the producer chooses the videotape or film format that will be used. From that point, equipment selections are a function of budget and inventory. (Video and audio equipment choices are too myriad to fit within the scope of this chapter.)

It is also the responsibility of the producer or technical director to secure the necessary equipment and an adequate supply of recording stock. If rental equipment or for hire services are involved, the further in advance plans can be made and equipment reserved, the better. If considering renting, establish a credit line with any rental facility. Few rental facilities can process a credit application on short notice.

Camera Support

Most studio cameras are tethered to an unwieldy supporting pedestal. The primary benefit of ENG and EFP is that cameras and recorders are mobile. More recently, lipstick style cameras can provide even more unique mobility and point-of-view options. For standard ENG and EFP cameras and camcorders, jibs, dollies, cranes, Tyler™ mounts, Steadicams™, helicopters and blimps are but a few of the many devices available to support a camera.[5]

Typically, the director determines camera angles. While a standard tripod and fluid head may be seem practical in most situations, directors should consider the options and advantages other support systems may offer.

Audio Concerns

Two common audio problems are dead mics and hum. Thanks to the popularity of the condenser mic, dead mic batteries have become a risk, particularly if batteries remain in mics for long periods of time. At best, replace the batteries before a production begins. At the least, thoroughly check the operation of all mics and always have extra mic batteries on hand. (See Chapter 3.3: *Microphones* for details on microphone selection and techniques.)

Hum can be a more difficult problem to resolve. In addition to the potential for ground loops, the production in the field introduces many variables into the audio and ac wiring equation not found in studios. Grounding problems are most likely. Check all audio circuits for hum and use extreme care to ensure that there is no ac leakage to the audio grounding system.

Not all field locations are ideal for all types of mics. Lav mics are popular for field work because they are virtually invisible, and they automatically follow the speaker. They also tend to eliminate background noise due to their close proximity to the body. However,

[5] Avgerakis, George, "You Want Me to Shoot Where?" *Videography Magazine*, Miller Freeman P.S.N. Inc, New York, Volume 20 Number 6, pp 40–42, 44–46, 112 June 1995.

according to some audio experts, lavs tend to muffle the sound as compared to a boom mic. Most experts agree that mics mounted on booms are preferable because they sound better and they can be controlled.[6]

Wind is another common source of problems on location. A good baffle or windsock will reduce or eliminate wind noise. Windsocks are made from a variety of materials. Typically fur works best, and thicker fur works better.

Lighting

The element of control separates ENG from EFP. In EFP, the ability to manipulate the lights, camera and action is the source of creative control. Lighting is an art and well beyond the scope of this chapter. See Chapter 5.17: *Lighting for Television* for more information on lighting.

The Remote Bag

Most people who work in remote production for any length of time build a remote bag. A typical remote bag contains a variety of disaster averting tools and useful gadgets.

Items found in many remote bags can range from gaffer's tape and permanent ink markers to volt-ohm-meters and portable test generators. The bag usually also includes various audio and video cable adapters, bullets and turn arounds that can come in handy.

Typically, people who shoot ENG rely on their remote bags for quick access to supplies that may be difficult or impossible to find on location.

The use of three-way to two-way electrical outlet adapters is not recommended. These adapters, commonly called suicide plugs did not gain the nickname by accident.

Other items include extra audio and video cables, chewing gum, work gloves and pre-packaged wet hand wipe towels.

Shooting Styles

ENG has evolved into a shooting style, characterized by less-than-perfect framing and heavy use of ambient light. It tends to evoke a sense of immediacy and importance. For single camera productions, style is limited to moving the camera to change the angle. The change may be live or edited in later, such as the addition of cutaway shots.

Far more popular, when possible, are multiple camera shoots. The video outputs of multiple cameras can be synchronized and switched live on a production switcher for live feeds. Alternatively, cameras can be synchronized with time code and edited together in post. With proper planning, either method can yield equal results.[7]

A newer shooting style is evolving as digital video editing gains momentum. People are learning that there are special considerations when shooting for non-linear postproduction. One common misconception is that non-linear editing does not require offline editing. In fact, digitizing video from field videotapes is a real time process (or 4× speed with some digital tape formats). Also, the limited capacity of most hard drives and disk arrays will not allow much space for unused clips.

People shooting for non-linear editing are finding that accurate time code logging during the shoot can save a great deal of time in post. See Chapter 5.9: *Video Editing Systems* for more details on shooting for the non-linear edit.

The Day of the Shoot

Whether the shoot is an emergency call to the scene of a breaking news event or a project that is well planned in advance, the basic procedure is the same: Get out the checklist.

With checklists developed for all categories of equipment, video, audio, lights and communication, inventory all items necessary for the shoot. Verify every item, regardless of size, including all cables and connectors. Using the lists, there is no excuse to ever show up for a shoot without all the pieces to make the shoot work.

For most EFP projects, setup may take a few hours or a few days, depending on the scope of the project. Take care to leave enough time for the crew to setup, including allowing time for potential repairs.

Be sure all crew members know their call times. It is best to provide each crew member with a personalized memo, complete with call times, maps and any other information that the crew should be aware of.

Upon arrival on the scene, check for electrical and other hazards first. Then, follow the plan.

SPECIFYING A REMOTE PRODUCTION VEHICLE

The market for remote production vehicles (RPVs) is a limited one, which forces the cost of such a vehicle to be high. Errors in specifying the vehicle can prove very expensive. Avoiding buyer's remorse, endless criticism, inconvenience, and potential trouble, the best approach to shopping for a production vehicle means asking the right questions; because unlike other TV plant projects, RPVs cannot be planned for expansion. Once a basic chassis is ordered, a number of aspects of the RPV become fixed.

Levels of Production

There are several levels of production requirements that will determine the overall scope and subsequent price tag of the project. On-site commercial videotaping requires the least from an RPV; live professional sports coverage can stretch the capabilities of the best equipped system. A good deal of the complexity of a vehicle is reduced if the vehicle is used only for remote

---

[6] Grotticelli, Michael, "Location Audio: Boom or Bust", *TV Broadcast Magazine*, Miller Freeman P.S.N. Inc, New York, pp 28, 62. May 1996.

[7] Reizner, Dick, "Be Prepared" *AV Video Magazine*, Montage Publishing, Inc., White Plains, NY, Volume 16, Number 12, pp 102, 104, 106–108, 110–111, December 1994.

videotaping. Such an operation leaves all editing, graphics insertion, effects system, audio sweetening, and other post production tasks in the studio. When the vehicle becomes the source of finished productions for live telecasting, the complexity and cost increase considerably because of the extra equipment and personnel requirements.

Single-Unit Vehicles

Once the RPV use is identified, an initial budget can be developed and the search for a vehicle type started. The single unit vehicle (with cab and box on a common chassis) is more maneuverable, easier to drive and park, and far less expensive than a tractor trailer system. For many stations, the step beyond the typical ENG vehicle is a cutaway van chassis with a cube-shaped box attached to the frame directly behind the driver. The height of the box offers adequate headroom and is commonly about 14 ft long with a 7 ft width. Approximately 100 square feet of useable area is available for work and storage space.

Single unit vehicles have drawbacks. One is a 29 ft maximum box length and resulting storage space limitations. Because of the drive shaft connecting the engine to the rear driving wheels, the under the truck storage is reduced. Belly box storage depth is usually limited to about a three feet, which prohibits its use for long tripods, light kits, cables, and bulkier items.

It may be instructive to contact a local truck dealer for the cost of unmodified vehicles of various sizes. The addition of two or three cameras, one or two VTRs, control equipment, an on-board generator, and possibly a microwave mast and transmitter electronics quickly raise the price. On the plus side, this unit will function well as a backup ENG vehicle when it is not involved in other field production assignments.

For somewhat more room, consider a stretch van, recreational, ambulance type vehicle or straight truck, each of which may be outfitted as a very serviceable RPV for two- to four-camera productions with on-board graphics. A straight truck, with a 29 ft box provides approximately 200 square ft of deck space. Systems based on straight trucks are often used for multicamera sports events, with four cameras, slow motion VTRs, graphics and a variety of sources. As a rule, they do not include a generator or microwave mast. If shore (on-site) power is not available, a generator can be rented locally.

The weight load and its distribution in the vehicle may require attention. A large concentration of weight over or behind the rear axle may create problems in steering. Ideally, the greatest concentration of weight should be fairly far forward. However, if the weight between the front and rear axles is too great, a tandem axle may be needed to relieve some of the strain. Such an axle must be ordered before the vehicle is constructed.

RPV builders may include a tag along axle behind or a pusher axle in front of the original rear drive axle. The tag along is particularly useful in stabilizing the vehicle, but does nothing in regard to weight distributions. A pusher axle improves load distribution and adds stability, while also reducing chassis flex. Modifications to the suspension may include heavier springs and shock absorbers, as well as electrical or hydraulic stabilizers.

The vehicle chassis size is directly related to the number of cameras the unit will be expected to handle. For a complex unit such as a complete 40 ft semi-network type vehicle including multiple cameras, switching and distribution equipment, recorders, titling and effects packages, expect a multimillion dollar price tag.

Interior Designs

The layout of an RPV interior does not provide for a great deal of flexibility. It would be wise to examine some existing vehicles to see what types of layouts might fit the expected requirements. It may also be a worthwhile investment in time to construct an actual-size cardboard mockup of an RPV interior, including control panels, rack layouts, and monitors. Such a mockup is an inexpensive engineering tool to get feedback from the staff and management before holes are drilled and wires are cut. In working with actual-size models, the staff will get a better feel of the proposed environment. Correcting a mistake in cardboard is much easier and less expensive than fixing an error in sheet metal and angle iron in the real vehicle. Combined with a logical, documented analysis of needs and wants, the mockup will help to prepare the purchase order.

While not as realistic as the life-size approach, miniature cardboard models can be used to check the fit of equipment into a floor plan. How well this works depends upon the accuracy with which the models are scaled. Computer assisted design (CAD) software offers another method by which various arrangements can be checked for proper fit. The basic floor plan is laid out in one layer of the drawing, a second drawing layer is used to locate racks and control equipment. If each element of the proposed design is created as a separate object, it is a relatively easy procedure to manipulate the positioning of the individual objects.

If an error in judgment or a lack of organized input to the builder results in alterations that must be made to the finished RPV, the minimum cost of the changes will be nearly threefold compared to the cost of doing it right the first time.

Equipment and people working in an enclosed area operate more reliably with a comfortable environment. When the truck is parked in an asphalt lot outside the stadium on a hot Sunday afternoon, heat buildup inside the vehicle can be considerable. Proper insulation and an air conditioning system designed for the heat load will keep the staff and the equipment at their best performance level. Formulas are available in texts and handbooks on heating and air conditioning for calculating the optimum capacity for an air conditioning system, given the wattage rating of the equipment, the number of people that will be in the area, insulation rating, outside temperature, and other factors.

MAKING IT MOBILE

In the case of the large production van, some arrangement must be considered to transport the vehicle from place to place. The cost of leasing a semitractor varies with the length of the lease period, with long term contracts being more economical. The lease plan usually includes normal and emergency maintenance and repairs, licensing, and taxes required for the tractor.

To purchase a semitractor, plan on a price tag over \$100,000 for a new unit. A serviceable, used tractor will probably carry a cost about half that of the new one, depending upon age and condition. Remember to add the weight of the trailer and all equipment and supplies when specifying the size and load capability of the tractor. If the vehicle will be used over a wide area, considerations may include a tractor with sleeping berth for long haul trips. Keep in mind, however, that reliability is a key. If the production system is to be in position at a specified time on a given date, it must be there!

An additional expense in the operation of a semitractor is the driver. Reliability and experience are very important. Requirements for drivers of network vehicles include a spotless driving record and experience with all types of driving conditions in all four seasons and in all areas of the country. Because the driver is in charge of a multimillion dollar package of equipment, a very professional attitude toward the job is important.

The Department of Transportation also has requirements for drivers of vehicles with a gross vehicular weight of more than 10,000 pounds. DOT regulations involve a record of duty status report to be kept by the driver. (The records to be kept may be found in DOT regulation Title 49 CFR, Section 395.8.) Requests by DOT inspectors to inspect the record could result in substantial monetary fines and a forced out-of-service period for a driver if rules have been violated. In some areas, additional local requirements more strict than those of the Federal Government will also apply. The driver must know all such rules for areas where the truck will travel and must meet medical physical requirements specified by the DOT.

Alternatives: To Lease or Buy

If there are uncertainties about the field production activities or the amount of expected use is limited, leasing presents a practical alternative. With a leased vehicle, production concepts can be checked out. Feasibility of remote activities for the station can be proven. Then, if the station later determines that the purchase of a vehicle is the proper step, valuable experience will have been gained toward specifying the purchase.

If the station owns its vehicle, it is responsible for continuing equipment maintenance, vehicle upkeep, and financing. Any time the system is out of service, it is a financial liability, so the need to keep all parts operating is essential. Under a leasing program, most major maintenance is handled by the leasing organization. The availability of maintenance facilities and the maintenance program offered should be important criteria in the selection of a leasing company.

The cost of leasing will vary with the equipment package desired, the size of the system, and with the area of the country. Many of the plans arrive at a total from a per item equipment schedule. The agreement can be written to include a complete operating crew for the required number of hours per day.

The purchase of a turnkey production truck is expensive, even for the simple ENG super van configuration. If the purpose is strictly ENG, perhaps a four-wheeled vehicle would provide the solution. With some protective storage for the camera, recorder and other essentials, and a microwave system, there should still be enough room for a camera operator, microwave technician and the reporter. Such a project might be easily handled by members of the engineering staff.

As the size of the vehicle increases, so does the complexity of the installation and the significance of the decision whether to do it yourself or buy a turnkey system. Outside manufacturers usually build trucks following a general design that reflects their experience and user feedback. Input from experienced station personnel will be valuable for whomever builds the vehicle and helps to determine a configuration that maximizes the vehicle's benefit, usefulness, and practicality, while remaining satisfying to the buyer.

If the plans include live remotes, the production vehicle will probably require microwave equipment to feed signals back to the studio. In addition, some plan should be made for two-way communications for cueing purposes. Fortunately, there is a range of choices, including cellular telephones, subcarrier transmissions, and mobile radio. As the distance between the remote site and the station increases, the communications selections could include a channel on the microwave system or even a connection through a mobile satellite facility. Another approach to get the program back to the station could use an established common carrier network.

In most large metropolitan areas today, the use of RF communications channels, particularly in the microwave bands, also means dealing with problems of interference. An extensive frequency coordination program has been instituted by the Society of Broadcast Engineers. Chapter 1.5: *Frequency Coordination* provides an in depth discussion of the problems and procedures of frequency coordination.

Lease Out: An Aid to Financing

The status of owning a vehicle and the convenience of not having to arrange for a production truck every time one is needed are plus factors. If the station's production schedule does not keep the system constantly in use, then it could be leased to others as a means to help in financing the purchase. Planning for this arrangement should include answering some additional questions.

1. Will the equipment complement make the truck desirable to others?

2. Given the station's planned uses of the vehicle, what is the likelihood that the station might need the vehicle when someone else is using it?
3. Will the vehicle need a staff when on lease out assignments? Depending upon the experience of the lessee, it may be advisable for the owner-lessor to provide some experienced staff to help ensure safe and effective operation.
4. How much effort would be required of the station's staff to prepare the vehicle for use by others and how much effort would be required to service and test the systems when the vehicle is returned?

PLANNING THE EQUIPMENT PACKAGE

Shopping for remote production equipment is as critical as buying for the studio. The right purchase includes several specific criteria:

- Quality
- Reliability
- Operational Simplicity
- Maintainability
- Compatibility with other equipment used by the station
- Price

Often equipment suppliers make special package offers, which provide significant cost savings over the purchase of individual items. While a program of getting bids on equipment packages takes time and may involve a good deal of paperwork, the process can result in a price tag that will allow an extra or two to be included.

The following sections include points to consider in various equipment areas. Most of the comments are directed toward portability. If a facility has already had experience with field production equipment, that experience will no doubt be a primary factor in future equipment purchases. Past experience could also be a reason to use an entirely different line of thinking. Applications of the decision spreadsheet can prove quite useful in this stage of system planning.

Acquisition Equipment

If the station is already active in ENG, has a major investment in the equipment of a single format, and has experienced good results with that equipment, then the purchases for additional field production units should stay with the same format. Where several systems are in use, the greater the similarity among the cameras and recorders, the easier it will be for operators to switch from one system to another, as the need arises. The more familiar an operator is with the equipment, the greater the effort the operator can put toward visual concerns. The performance advantage of a new format may be outweighed by such practical operational concerns.

An initial evaluation of equipment, based on published specifications, is another opportunity to use a spreadsheet approach to compare information about different models. List the models as column heads, let the rows be the more important specifications. Numbers can be filled in from manufacturers' literature. Table 5.17-1 is an example of items of interest for cameras.

Field Cameras

Selection of a portable camera for ENG has been simplified by improvements in charge coupled device (CCD) technology.

Two areas are left open for discussion: the type of output and the lens. Many of the current generation of cameras are designed to mate with a series of adapters. The adapter determines the output signal format: composite with multicore, coax, triax, or fiber optic cable; luminance and color components; Y/C components; and adaptations for docking recorders.

The trend in lens designs meets the special characteristic of CCD cameras. Because CCDs are cemented to the beam splitting prism and cannot be moved individually for fine focus, as was the case with tubes, the lens system must be optimized to compensate for chromatic aberration. While an uncompensated error may be visible as a slight fringing of color, lenses designed for CCDs bring all three color images into focus at the same plane.

Lens length may be partially based on a matter of personal choice and partially on the type of event to be covered. Lenses with a greater zoom ratio permit greater flexibility in camera location for coverage of some events, but one characteristic of long lenses should be kept in mind: the longer the lens focal length, the more pronounced are even minor movements. As lens focal length increases, the need for a more stable support than the operator' shoulder may be required.

Table 5.17-1
Spreadsheet to Compare NTSC Cameras.

Attribute	Model 1	Model 2	Model 3
CCD type & pixel array			
Sensitivity & max.sens.			
Resolution depth of modulation S/N Ratio			
Output format			
Adaptors: NTSC composite RGB<R>Y/R-Y/B-Y R/C(S-VHS)			
Cables: Multicore, Triaxial Coaxial, Fiber optic			
Docking VTR			
Control units: RCU, LCU, CCU Multicamera CU			
Camera Head Weight			

Another factor in camera selection may be of importance particularly to camera operators who rove through crowds or shoot video at sports activities. Several special shapes of the camera package are designed for better operator vision. Standard camera designs create a blind spot to the right of the operator. A more ergonomic design reduces the height of the camera body in such a way that the operator can see toward the right while shooting. Although the appearance may seem odd at first, many operators come to appreciate the design feature after they are acquainted with it.

Operators should be aware of the relative levels of illumination in which they are shooting. Most CCD cameras can produce usable images to quite low levels. The use of a shutter with CCDs has become common place to avoid smearing in the pictures as a result of highlights or difficult contrast conditions. Integral neutral density filtering and an automatic iris are important in coping with widely differing or rapidly changing levels of light. For reference, Table 5.17-2 notes several common lighting conditions and the relative light levels that may be encountered:

Nearly all portable cameras include two or more levels of color correction filters to assist in color balancing for various light sources. These filters, along with color balance memories, can be used to produce realistic color in the output pictures.

Field Recording

Field recording formats include analog component and digital formats. Making a choice among these presents a real challenge, because all of the professional systems perform quite well. Beyond the problem of compatibility with existing equipment, consider how the product will be used.

ENG cameras typically provide for microphones to be mounted directly on the camera body with a connector tied into a program audio circuit. The purpose of the video acquisition will decide the type of microphone pattern to be used. For shoots that involve general audio coverage, an omnidirectional or general cardioid pattern will serve nicely. If it is necessary to pickup only specific talent on the audio track, a shotgun unidirectional instrument will prove a better choice. This feature can avoid the problems of microphone cables at the remote site. Many instances will still require the talent to use a handheld microphone for best vocal pickups. With a miniaturized wireless microphone receiver attached to the camera, a wireless handheld or body pack microphone solves the cable problem. Wireless operation does have the drawback of possible interference from external sources.

When multiple microphones are used, a need for some audio mixing capability exists. If the portable recorder includes multiple audio channels, mixing may be deferred until post production. Much more freedom and creativity is available during post production work when separate audio tracks are recorded.

Table 5.17-2
Light levels of Common Lighting Conditions.

Full sun	100,000 lux
Overcast at noon	10,000 lux
Interior by wind	1,000 lux
Interior work area	100 lux
Full moonlight	0.2 lux

Cables and Links

More a concern for EFP and OB projects is the question of getting the signal back to the truck. For cameras that operate on the various types of cables, the choice must be made between multicore, coax, triax and fiber cables. Fiber optic options permit the greatest distance between the camera and the production vehicle of the four types. If distances are greater than those suggested for a single fiber optic cable run, relay units are available to double the length. During the construction of some stadium facilities, the needs for television are often included in the design. Cabling may be permanently installed to various locations in the stadium, reducing the setup time required for the TV crew.

Cableless cameras may be outfitted with a miniature microwave system to beam the output to the central production center. These systems may include a tracking antenna, which permits a return channel for communications between the truck and the operator. If the line-of-sight path is blocked, a relay becomes necessary. Such a relay in an open stadium setting might use an antenna (and electronics) held aloft by a balloon attached to the upper, outer wall of the stadium. The relay unit remains relatively fixed positioned with respect to the truck parked outside the stadium, while the tracking system in the camera unit keeps the primary transmitted energy aimed at the relay antenna.

GETTING BACK TO HEADQUARTERS

Live remotes bring together every complexity of the television broadcast industry. Not only must the remote truck operate perfectly, so must the program link from the remote originating site to the base station. Technology has produced a variety of methods for transporting the program signal back to the station. Options include wired connections (via telephone company facilities), microwave video radio (ENG), and satellite relays. The distance between the production site and the station can be a deciding factor in the transmission choice.

Terrestrial Microwave

If the site is local, for example, the most expedient link between the two locations is probably by microwave, especially if line-of-site operation is possible. If obstacles to a direct line path exist, perhaps the use of a reflector can redirect the signal around the obstacle. Microwave system design is detailed in Chapter 6.1: *Microwave and STL Systems.*

Although microwave frequencies may already be assigned in a metropolitan area, a check on frequency coordination is advisable to make certain that interference can be avoided. Often cooperation between stations is needed to ensure interference-free operation for special events. A case in point is the renowned July Fourth celebration in Boston, during which a live performance of the Boston Pops Orchestra is telecast nationwide by PBS. In addition to the site-to-station channels, local site microwave between boats, helicopters, nearby buildings, and the production truck are in use. Because of a limited number of frequencies, the production is made possible only through cooperation between the various local stations in frequency coordination and equipment sharing. Coordination activities for such scheduled events typically begin months in advance.

As distances between the remote site and base facilities increase, microwave by common carrier services becomes practical. Demands for more communications channels have lead to the use of digital transmissions to accommodate multiplexing as well as bandwidth reduction through data compression. Several services are now offered using the technology of digital transmissions.

Satellite Backhaul

Program relays by satellite permits the remote site to be almost anywhere in the world. Many of the relay and backhaul services available by satellite also involve digital signal carriage with data compression to help deal with the amount of traffic that must be transmitted at all times. Depending upon the destination for the signal, the satellite system may be the more practical route to be used. In larger urban areas, occasional use satellite time can probably be obtained on short notice and is competitively in price with fiber.

For remote operation, the development of smaller Ku-band terminals and antennas offers several advantages. The size of a C-band antenna was too large to easily include an uplink antenna as a part of a remote production vehicle. As a result, such C-band equipment was often taken along to the remote location on a separate vehicle. Ku-band uplink equipment can actually be mounted on a production vehicle and immediately deployed upon arrival of the truck at the site. An option provides the convenience of almost completely automatic satellite antenna alignment through the GPS navigation system. The engineer needs only to key into the control computer both the geographical coordinates of the site and the satellite through which the relay will pass. Within minutes the system is setup and operational.

Fly-away C-band and Ku-band equipment is available with folding dishes in the two meter diameter range. Electronics, antenna, and mount can fit into a half dozen or so suitcases that can be checked as airline luggage. Equipment manufacturers and the operators of satellites and commercial teleports can provide technical and cost information on which to base planning decisions for the use of such facilities. With the growth of switched fiber networks, the available combinations of satellite, fiber, and terrestrial microwave links offer almost unlimited possibilities for live remote production anywhere in the world.

COMMUNICATIONS

Keeping in touch with the remote production crew is one of the secrets of a successful production session. Even for a simple ENG assignment, where only a camera operator is dispatched to gather local color for a special closing segment, communications between the operator and the station should not be forgotten. It may be necessary to direct the operator to another assignment before returning to the station or the operator may require assistance from the station before returning. As the production project becomes more involved, communications becomes even more important. When ENG goes live, cues to the technical staff and talent are as crucial as those to a remote team located halfway around the world.

Options now available for keeping in touch with mobilized members of the staff are widely varied. Cellular telephones and two-way radio provide convenient bidirectional links between the control room and the remote van over a wide area of operation. Both require licensing. Cellular telephones can also be used to transmit data between two points, including facsimile type data transmissions. Just as facsimile messages have become popular throughout modern business, the remote copier capability is particularly valuable in sending graphic type material that would require much more complex equipment for handling.

When the remote production crew leaves the area served by the TV station, two-way communication with the station can be carried along with program signals by microwave or satellite through an order wire subchannel of the existing connection. The roaming feature of cellular telephones can keep a remote crew in almost constant contact from most areas of the United States and Canada. Undoubtedly this will expand to a worldwide connection, as satellites play a greater role in linking terrestrial communications facilities.

Communications among the staff at the remote site may mean another challenge for engineering. Operators of wired cameras remain in touch through the intercom circuit available with all cameras. Roving, nonwired camera operators are generally equipped with two-way radios or wireless intercom equipment. In a facility such as a stadium, where similar communications are required on a frequent basis, numerous wired systems may be permanently installed and staff members are required only to be plugged into the network.

These are only a few of the equipment areas that must be carefully considered before the purchase order for a remote production vehicle and equipment package can be finalized. You will also need to consider video switchers, audio mixers, on-board audio and video recording equipment, signal distribution and monitoring products. No less important to the comfort and efficiency of the operations staff are such concerns as the furnishings in the vehicle.

KEY TERMS

Cutaway. A particular type of scene, usually shot after the fact, to be edited in later. The scene is either a person talking (usually a shot of the back of the head), or a person listening (usually physically responding to what the other person is saying). Cutaways are used as a video source to cover the edits of incongruous scenes.

Shore power. A source of electricity other than a portable generator. Typically, shore power is electrical service from the local electric utility.

BIBLIOGRAPHY

Balonis, Ronald, "Planning an STL System," pp. 116–138, October 1988.

Dick, Brad,"Satellite Uplink Trucks," *Broadcast Engineering,* pp. 44–62, January 1986.

Dick, Brad, "Applying Cellular Technology," *Broadcast Engineering,* pp. 36–44, January 1988.

Hardy, Steven D, "Taking to the Skies," *Broadcast Engineering,* pp. 76–80, January 1987.

Lehtinen, Rick, "Making the Most of Remotes," *Broadcast Engineering*, pp. 38–48, 99, January 1990.

Lehtinen, Rick, "Communicating with the Field," *Broadcast Engineering,* pp. 48–60, January 1991.

Pizzi, Skip, "Remotes Revisited," *Broadcast Engineering,* pp. 26–46, January 1991.

Renwanz, Karl, "Building an ENG Network," *Broadcast Engineering,* pp. 26–42, January 1986.

Smith, Ron, "Catching the News Crew," *Broadcast Engineering,* pp. 46–56, January 1988.

Soseman, Ned, "Planning A Remote Production Vehicle," *Broadcast Engineering*, pp. 68–74, January 1987.

Tourkow, Bob, "Intercom System Design," *Broadcast Engineering*, pp. 74–96, March 1986.

Wolf, Jan, "Mobile Mast Safety," *Broadcast Engineering,* pp. 26–32, 1988.

Equipment Reference Manual, (See most recent annual issue).

5.18
PRODUCTION CONSIDERATIONS FOR DTV

DAVID LEATHERS
DIGITAL MEDIA LAB, CULVER CITY, CA

INTRODUCTION

At this writing, any discussion of the production realities of digital television (DTV) is a bit like shooting at a moving target. The only certainty is that it will be a rapidly changing environment that will present both opportunities and problems that have not yet been anticipated.

What forms DTV's implementation will take is still a matter of great speculation. The various political, market, technical and creative forces that will shape DTV's future are still evolving. The standards and uses that will eventually emerge will not be defined until broadcasters, manufacturers and content creators have gained real experience in the use of the medium. However, it is reasonable to assume that there will be a fairly wide range of approaches, both initially, and as more technological resources are introduced.

PROLIFERATION OF FORMATS

In the earlier conceptual stages, DTV was thought of primarily as a delivery system for a digital form of high definition television (HDTV). However, as DTV moves from the theoretical to the practical, it is clear that high bandwidth HDTV represents only one use of the technology. There are currently 18 video format standards that have been adopted for DTV transmission at greater than SMPTE 259M resolution. However, the use of DTV as a high definition medium has not been mandated by the FCC. In addition to HDTV broadcasting, broadcasters are free to use the digital spectrum for multicasting multiple channels of standard definition video, lower resolution video, interactive programming, internet downloads or other digital forms to be determined. Consequently, the range of possibilities for content has expanded dramatically beyond high bandwidth HDTV.

Production models used in other entertainment areas, such as electronic games, the internet, DVD and video CDs and visual effects production, are useful in predicting the directions that DTV may take and the consequent production requirements and considerations. Each of these industries is highly responsive to rapid changes in their respective technological and market environments. Also, the costs and capabilities of production systems change rapidly, creating a volatile arena for technicians, producers and artists.

The net result is, in addition to all the other considerations that go into developing production properties, producers will have to consider a far wider range of production and distribution formats than ever before.

HDTV PRODUCTION

Forms of high resolution television have been displayed at NAB conventions for nearly a decade. In that time, the specifications have changed and multiple formats and standards have evolved. HDTV production equipment is just now becoming readily available, yet production in the HDTV domain has been limited to date.

There are several reasons for this. Since, until now, the equipment itself has been produced in small numbers, the cost-per-unit has remained high. Since, there is no installed base of display equipment, those productions that have been done in HDTV have had limited exhibition possibilities or have been converted to other formats.

Until more equipment exists, it is likely that most feature entertainment productions will be produced on film.

However, film cannot be produced live. The need for live HDTV content will begin to drive the need for pervasive affordable HDTV production systems. Accurate predictions of how long this will take are impossible at this time.

Film Production and Media Assets

In today's distribution environment, feature films are usually shot in 35 mm film. After the film is distributed theatrically, using 35 mm prints, it is usually telecined to standard NTSC, PAL and/or other video standards and distributed on broadcast quality digital and/or analog tape for broadcast, satellite and cable delivery. It is subsequently duplicated to VHS tape for distribution in the home video rental and other consumer markets. Other forms of video entertainment and information, which are not intended for film release, are often produced on video. The exception is when productions, intended for television or video, are shot in film, in anticipation of a possible future re-release in a higher resolution format. This is a method employed to protect the future value of unique and/or valuable content, or media assets.

35 mm film specifications exceed the resolution and contrast of all forms of HDTV by a significant margin. For this reason, major television productions, as well as feature films, have been shot in film for many years. It is still the best way to ensure that the material shot

in the present, will retain its full media asset value in the foreseeable future. Once the media assets are in this high resolution form, they can be telecined and down converted to whatever lower resolution format is required for delivery at that time. As delivery systems improve, the original media assets can be copied into better formats, retaining their original value.

Predictable Production Resources

There are other inertial factors that will likely keep high end production in the film domain for some time to come. 35 mm production equipment and experienced crews are widely available. Production techniques are well developed and understood. Production schedules and costs are generally predictable.

Also, in the last decade, the visual effects industry has developed many ways to work with film originated imagery in the digital domain. This makes it possible to create almost any imaginable visual effect at film resolutions that exceed today's specifications for HDTV. When it does become practical to produce feature entertainment in HDTV, the visual effects components will already be in place.

Visual Effects and Electronic Production

With the release of Pixar's *Toy Story* in 1996, audiences saw the first feature film that was produced completely in the digital domain. Similar end-to-end computer generated production of HDTV content is therefore possible. Since the beginning of production for *Toy Story*, which reportedly took five years to produce, advances in hardware and software have reduced the cost and production time required to levels that are feasible for television production. In fact, many of the facilities that were once dedicated to film visual effects production are now also involved in the production of graphics, animation and visual effects for lower resolution media, such as television and electronic games, in addition to their high resolution film activities.

The primary tools of visual effects production are computer hardware and software. As the toolset has evolved away from dedicated hardware systems (black boxes) and towards open computer platform based systems, visual effects production has reached a point where it is essentially resolution independent. Once again, the resolution requirements for digital film production exceed the requirements of all anticipated HDTV formats. This means that essentially all available production tools and techniques in use today for film visual effects production, are, or will be available for HDTV production, regardless of the resolution or aspect ratio.

Live HDTV Broadcasting

The major disadvantage to both film production and computer based production is that they can not be done in real-time. Television style production of live events is unique to the video domain. HDTV production systems will have to be used for HDTV presentations of live events.

The engine that will drive the development of HDTV production systems will most likely be this need for live broadcasting. Sports production is likely to play a large role in this development. Major sporting events are frequent enough and have the budgets to push the technical envelope in live production. Olympic competitions have already served as some of the first ventures in HDTV production.

HDTV Formats, Progressive Scan and 16:9 Aspect Ratios

While video professionals have been evaluating HDTV technology and picture quality using studio quality equipment and monitors for years, the quality of the initial consumer installed base of HD capable home viewing equipment will be subject to significant trade-offs in picture quality in deference to price.

While, tests continue, there is widespread *speculation* that the differences in the aspect ratio between 4:3 displays and 16:9 displays and the differences between interlaced and non-interlaced (progressive scan) displays are far more dramatic than the effects of increased pixel density.

The Computer Environment

As the shift to DTV takes place, perhaps the largest consideration affecting those involved in production, will be the central position that computers will take in the entire process, from preproduction through delivery to the home.

Digital videotape formats and serial digital systems in production environments have improved on the traditional production process but have not fundamentally changed the production process. Digital tape machines and switchers are used in much the same way as their analog predecessors and are used together in systems with the older analog equipment.

DTV's demand for both compression and interactivity will drive more of the production process into the computer domain. This will fundamentally change the production process. The great strides being made in the production capabilities of computer systems in the last few years are indicators of a process that will be greatly accelerated by the implementation of DTV. For evidence of this, consider the current role of online nonlinear editing systems and computer based graphics production in television production and compare this to their roles and capabilities of only a few years ago.

While these roles have increased dramatically, they have still remained a part of a process with the purpose of putting standard resolution, linear programming on the air. The demands of DTV production will require the flexibility of more computer-centric systems to accommodate the wider range of output demands. These include multiple streams of video, compressed or uncompressed, at varying resolutions and aspect ratios, multiple streams of compressed digital audio, computer graphics and animation and interactive programming at the minimum. The integration of data services, interactivity and more advanced forms of video will further centralize the role of computers.

Software and Computer Programming

It follows that this greater reliance on computers in the production process will create a need to understand computers and computer programming. A look at today's visual effects and interactive gaming industries offers some insights into the role that software and computer programming will play in DTV production.

In the visual effects industry, film effects are generated using a variety of software packages in multiplatform production environments. Typically, a process will call for more than one software program to be used. Also, the interoperability between these software packages is not adequate for the producer's purpose and custom software must be developed to solve the incompatibility. Therefore, most visual effects facilities support a software development department to work with artists and solve production problems. It is reasonable to assume that this type of approach will be necessary for the leaders in DTV production. Like visual effects producers, DTV producers will be pushing the technical envelope, as well as the creative envelope. The only way to do this in a computer production environment is to have available skills to solve problems and offer solutions that involve custom computer programming solutions.

This will create high demand in production environments for computer programming skills and for producers, technicians and others involved in the creative process to understand the role of software in the overall production process.

Interactivity

While questions associated with HDTV, and multicasting are important, of at least equal importance are the questions posed by the interactive aspects of DTV. In terms of how DTV is used, interactivity may play a larger role in the future orientation of consumers to television than HDTV or the proliferation of programming choices offered by multicasting.

While the production of linear programming is familiar territory to film, television and video producers, broadcasters and distributors, the development of interactive programming and the interactive infrastructure that will become possible with DTV is new and largely unfamiliar.

Programming interactivity is inherently custom work. There is no standard for interactivity. This is because, in an interactive experience, the design and implementation of the interactivity is fundamental to the creative concept.

Interactivity must be programmed into a production, one step at a time. It is a labor intensive activity that requires specific training and skills.

As with other creative computer processes, such as graphics, animation and visual effects production, the more innovative the implementation, the greater the demand for custom programming. While it is possible to create interactive presentations, using off-the-shelf tools, the most dynamic results are achieved using custom programming techniques.

Interactive programming will be addressed differently by individual DTV producers and broadcasters. While some may perceive demand for a lot of interactive programming in their market and get involved in the production of programming to meet that demand, others will not pursue it as a staple, in part due to the costs and skills required for production. However, even those stations that do not broadcast interactive programming, may need to transmit data that can be used by the receiver to create interactive menus and navigation systems that allow the user to select their services.

Interactivity and Multiple Simultaneous Cameras

One of the ideas for the use of interactivity in DTV is the delivery of multiple camera angles to the viewer along with the viewer's ability to select between them. The viewer would see small proxies of several angles along the side or bottom of his screen, and flip between them to display his choice on the main screen area.

From the production standpoint, this means that multiple cameras have to be framed, focused and online at any point in time. In film style production done offline, this means that there will be more cameras shooting simultaneously, or more retakes from more angles. In live production, this means that camera operators will have less time to hunt for shots and frame and focus and will have to be live more of the time. In a large sports production, where there are a dozen or more cameras, directors will have to be sure that a minimum number of those cameras are online and providing usable angles at any point in time. The possibility of this kind of coverage is exciting from the viewing standpoint and undoubtedly will be tried as technology permits.

What this kind of interactivity means to the line producer, whether live or not, is that more source material will be required to fill the same amount of linear time.

Immersive Video

A future consideration, that admittedly is in the early stages of development, but has very important possibilities in the future of DTV is *immersive video.* Immersive video is the use of input from multiple cameras surrounding an area to produce a videographically generated 3D database of that area that is live and changes in real-time.

Once the 3D database is created (in real-time), an interactive viewer will be able to navigate the 3D environment using a virtual camera controlled by a joystick or similar device. Users will be able to walk through the video environment, just as video gamers can walk through the 3D environments in today's 3D computer generated game environments. For instance, the sports viewer will be able to watch their favorite game from a virtual camera position anywhere on the field.

CHARACTERISTICS OF THE VIDEO SIGNAL[1]

HDTV has improved on earlier techniques primarily by calling more fully upon the resources of human vision.[2] The primary objective of HDTV has been to enlarge the visual field occupied by the video image. This has called for larger, wider pictures that are intended to be viewed more closely than conventional video. To satisfy the viewer at this closer inspection, the HDTV image must possess proportionately finer detail and sharpness of outlines.

Critical Implications for the Viewer and Program Producer

In its search for a "new viewing experience," NHK conducted an extensive psychophysical research program in the early 1970s. A large number of attributes were studied. Non-technical people were exposed to a wide variety of electronic images, whose many parameters were then varied over a wide range. A definition of those imaging parameters was being sought, the aggregate of which would satisfy the average viewer that the television image portrayal produced an emotional stimulation similar to that of large screen film cinema experience.

Central to this effort was the pivotal fact that the image portrayed would be large—considerably larger than current NTSC television receivers. Some of the key definitions being sought by NHK were precisely how large, how wide, how much resolution and the optimum viewing distance of this new video image.

A substantial body of research gathered over the years has established that the average U.S. consumer views the television receiver from a distance of approximately seven picture heights. This translates to perhaps a 27 in. NTSC screen viewed from a distance of about 10 ft. At this viewing distance most of the NTSC artifacts are invisible, with perhaps the exception of cross color. Certainly the scanning lines are invisible. The luminance resolution is satisfactory on camera close-ups. A facial close-up on a modern, high performance 525-line NTSC receiver, viewed from a distance of 10 ft, is quite a realistic and pleasing portrayal. But the system quickly fails on many counts when dealing with more complex scene content.

Wide angle shots (such as jersey numbers on football players) represent one simple and familiar example. Television camera shooting, however, has long adapted to this inherent restriction of 525 NTSC—as witnessed by the continuous zooming-in for close-ups on most sporting events. The camera operator accommodates for the technical shortcomings of the conventional TV system and delivers an image that meets the capabilities of NTSC, PAL, and SECAM quite reasonably. There is however, a penalty, as illustrated in Figure 5.18-1. The average home viewer is presented with a very narrow angle of view—on the order of 10°. The video image has been rendered clean of many inherent disturbances by the 10 ft viewing distance, and made adequate in resolution by the action of the camera operator; but in the process the scene has become a small window. The now acceptable TV image pales in comparison with the sometimes awesome visual stimulation of the cinema. The primary limitation of conventional TV systems is, therefore, one of image size. A direct consequence is further limitation of image content; the angle of view is constantly constricted by the need to provide adequate resolution. There is significant, necessary and unseen intervention by the TV program director in the establishment of the image content that can be passed on to the home viewer with acceptable resolution.

Compared to the 525-line NTSC signal (or the marginally better PAL and SECAM systems), the Grand Alliance DTV system and the North American HDTV studio standard (SMPTE 240M) and its digital representation (SMPTE 274M) offer a vast increase in total information contained within the visual image. If all this information is portrayed on an appropriate HDTV studio monitor (currently available in 13, 19, 28, 38 in diagonal sizes) the dramatic technical superiority of HDTV over conventional technology can easily be seen. The additional visual information, coupled with

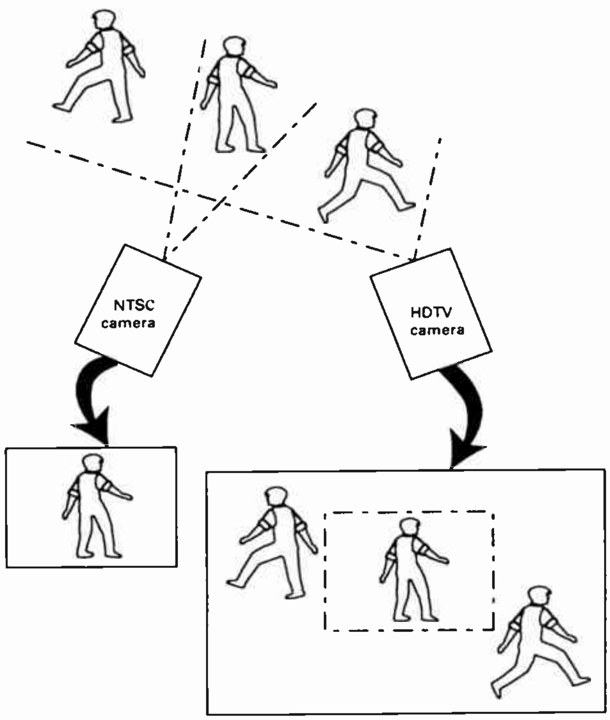


Figure 5.18-1. An illustration of the differences in the scene capture capabilities of conventional video and HDTV.[2]

[1] Whitaker, Jerry, *DTV The Revolution in Electronic Imaging*, McGraw-Hill, 1998, pp. 45-50.

[2] Thorpe, Laurence, J., "Applying High DefinitionTelevision," *Television Engineering Handbook*, rev.ed. K.B. Benson and J.C. Whitaker (eds.), McGraw-Hill, New York, p.23.4, 1991.

the elimination of composite video artifacts, portrays an image almost totally free (subjectively) of visible distortions, even when viewed at a close distance.

On a direct-view CRT monitor, HDTV displays a technically superb picture. The *information density* is high; the picture has a startling clarity. However, when viewed from a distance of approximately seven picture heights, it is virtually indistinguishable from a good NTSC portrayal. The wider aspect ratio is the most dramatic change in the viewing experience at normal viewing distances.

Image Size

If HDTV is to find a home for the consumer, it will find it in the living room. If consumers are to retain the average viewing distance of 10 ft, then the minimum image size required for an HDTV screen for the average definition of a totally new viewing experience is about a 75 in. diagonal. This represents an image area considerably in excess of present large 27 in. NTSC (and PAL/SECAM) television receivers. In fact, as indicated in Figure 5.18-2, the viewing geometry translates into a viewing angle close to 30°, and a distance of only three picture heights between the viewer and the HDTV screen. Compare this with the viewing angle for conventional systems at 10°, as shown in Figure 5.18-3.

HDTV Image Content

There is more to the enhanced viewing experience than merely elevating picture size. Unfortunately, this fundamental premise has been ignored in some audience surveys. The larger, artifact-free, imaging capability of HDTV allows a new image portrayal that capitalizes on the attributes of the larger screen. As mentioned previously, as long as the camera operator appropriately fills the 525 (or 625) scanning system, the resulting image (from a resolution viewpoint) is actually quite satisfactory on conventional systems. If, however, the same scene is shot with an 1125/60 HDTV camera, for example, and the angle of view of the lens is adjusted to portray the same resolution (in the picture center) as the 525 camera when capturing a close-up of an individual ball player on its 525 screen,

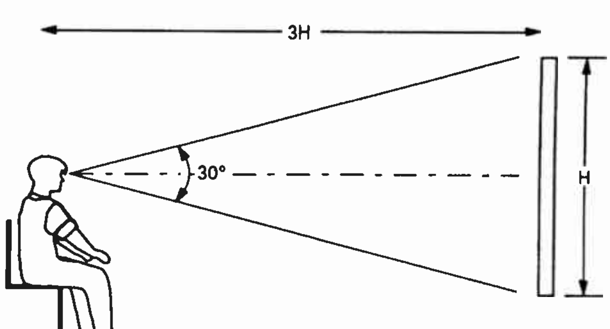


Figure 5.18-2. Viewing angle as a function of screen distance for HDTV.[2]

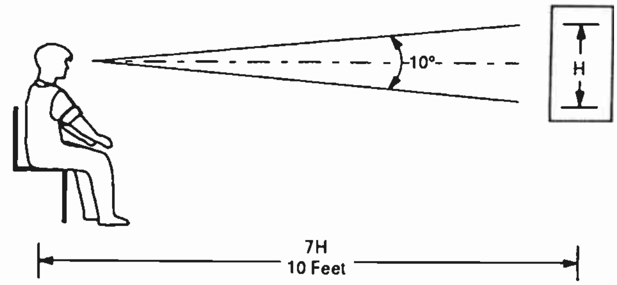


Figure 5.18-3. Viewing angle as a function of screen distance for conventional video systems.[2]

a vital difference between the two pictures emerges: the larger HDTV image contains considerably more information, as illustrated in Figure 5.18-1.

The HDTV picture shows more of the ball field—more players, more of the total action. The HDTV image is, thus, radically different than that of the NTSC portrayal. The individual players are portrayed with the same resolution on the retina—at the same viewing distance—but a totally different viewing experience is provided for the consumer. The essence of HDTV imaging is this greater sensation of reality.

The real dramatic impact of HDTV on the consumer will be realized only when two key ingredients are included:

- Presentation of an image size of approximately 75 in. diagonal (minimum)
- Presentation of image content that capitalizes on new camera freedom in formatting larger, wider, and more true to life angles of view.

Format Development

Established procedures in the program production community provide for the 4:3 aspect ratio of video productions and motion-picture films shot specifically for video distribution. This format convention has—by and large—been adopted by the computer industry for desktop computer systems.

In the staging of motion-picture films intended for theatrical distribution, generally no provision is made for the limitations of conventional video displays. Instead, the full screen, in wide aspect ratios—such as CinemaScope—is used by directors for maximum dramatic and sensory impact. Consequently, cropping of essential information may be encountered more often than not on the video screen. This problem is particularly acute in wide-screen features where cropping of the sides of the film frame is necessary in producing a print for video transmission. This objective is met in one of the following ways:

- *Letter-box* transmission with blank areas above and below the wide screen frame. Audiences in North

[3] Benson, K.B. and D.G. Fink., *HDTV: Advanced Television for the 1990s*, McGraw-Hill, New York. 1990.

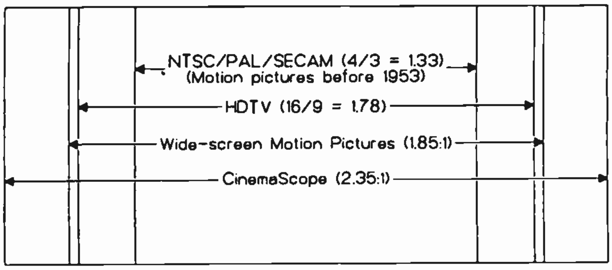


Figure 5.18-4. Comparison of the aspect ratios of television and motion pictures.[3]

America and Japan have generally not accepted this presentation format, primarily because of the reduced size of the picture images and the aesthetic distraction of the blank screen areas
- Printing the full frame height and crop equal portions of the left and right sides to provide a 4:3 aspect ratio. This process frequently is less than ideal because, depending upon the scene, important visual elements may be eliminated
- Programming the horizontal placement of a 4:3 aperture to follow the essential picture information (*pan and scan*). This process is used in producing a print or in making a film-to-tape transfer for video viewing. Editorial judgment is required to determine the scanning cues for horizontal positioning and, if panning is used, the rate of horizontal movement. This is an expensive and laborious procedure and, at best, compromises the artistic judgments made by the director and the cinematographer in staging and shooting, and by the film editor in post production.

These considerations are also of importance to the computer industry, which currently has keen interest in multimedia technology.

One of the reasons for moving to a 16:9 format is to take advantage of consumer acceptance of the 16:9 aspect ratio commonly found in motion picture films. Actually, however, motion pictures are produced in several formats, including:

- 4:3 (1.33)
- 2.35, used for 35 mm anamorphic CinemaScope film
- 2.2 in a 70 mm format

Still, however, the 16:9 aspect ratio is commonly supported by the motion picture industry. Figure 5.18-4 illustrates some of the more common aspect ratios.

CONCLUSIONS

DTV opens the doors to all kinds of new possibilities in programming formats for producers. This will be both a blessing and a curse.

The range of possibilities is so vast that producers will have more choices to make. Delivery format decisions will have to be considered at a fundamental level along with content. New skills will be required in the production process and producers will have to draw talent form the areas of visual effects production, internet programming, electronic games and computer software.

Funding projects will take on new dimensions. The great proliferation of potential channels, created thorough multicasting, will generate a need for more programming. At the same time, this proliferation of channels will dilute the audience, necessitating smaller production budgets for many projects.

As stated at the beginning of this chapter, how the implementation of DTV actually evolves will depend on a vast number of forces and interests. Producers need to increase their knowledge of as many of these areas as they can in order to be able to respond to the challenges that lie ahead and to take advantage of the many wonderful new capabilities that the technology will deliver.

Section 6: Television Transmission Facilities

Introduction by: Donald Wilkinson, Fisher Broadcasting, Seattle, WA

This is a time of remarkable changes in the television broadcasting industry. Many new players are competing with traditional broadcasters for the attention of the viewers. The quality of the picture that you transmit can easily make the difference between an enjoyable experience and driving the viewer away. With the conversion to digital television, even more care must be taken to preserve the integrity of your RF transmission path. Failure to do so can reduce your coverage area as surely as if you reduced power.

In this section, NAB has assembled a group of experts in television transmission facilities to address some of the issues facing broadcast engineers today. Much of the information contained herein will be very useful, especially for those of us being exposed to UHF transmitters for the first time. The chapters include a wide range of valuable information that will assist you in keeping your transmission facility operating properly. Subjects covered include microwave and fiber transmission, transmitters, combiners and diplexers, waveguide, antennas, measurements and satellite systems.

The section begins with the *Microwave and STL Systems* chapter, wherein Ernest M. Hickin and James H. Rooney III discuss the importance of microwave transmission and studio-to-transmitter links with regards to the broadcast engineer's responsibilities.

Gerald W. Collins and Robert J. Plonka discuss the relevant technology and provide the information needed for selection, installation, operation and maintenance of TV transmitters in today's environment in their *Television Transmitters* chapter.

Multichannel television sound is a generic term for the process of adding subcarriers to the aural carrier of an NTSC television station. In his *Multichannel Television Sound* chapter, Edmund A. Williams delves into the technical intricacies behind the generic term.

"Changes in telecommunications markets and technologies are challenging TV broadcasters to broaden the definition of their business and address new market opportunities." Neil Mitchell and NAB Staffer Kelly T. Williams explain how datacasting allows TV broadcasters to expand their business in *Television Data Broadcasting* chapter.

To help television broadcasters meet their compliance goals, John E. Leonard Jr. examines the regulatory requirement changes to Part 73 of the FCC's Rules and Regulations in the *Transmission System Control and Monitoring* chapter.

A waveguide is much more than a structure that guides electromagnetic waves. In his *Waveguide* chapter, *Handbook* Editor-in-Chief Jerry Whitaker discusses the several types of waveguides and its advantages over coax.

William Decormier's *Diplexers, Combiners and Filters* chapter reminds broadcasters that in order to optimize the performance of a station's transmission system, they must first understand and optimize the components upon which the system is built.

Thomas J. Vaughan's instructive *Antennas for Television Broadcasting* chapter begins with FCC considerations followed by a review of antenna design considerations as they relate to these demands and a review of specific applications. The chapter concludes with a look at current broadcast antennas and antenna systems.

Joseph W. Stielper and Ann Gallagher's *Television Field Strength Measurement (54–806 MHz)* chapter discusses planning a program of television broadcast field strength measurements and the analysis of the resulting data. FM broadcasters will find this chapter useful as well.

Signal transmission using fiber optic technology is the focus of C. Robert (Bob) Paulson's chapter on *Fiber Optic Transmission Systems for Broadcasting.*

As satellite technology for communications becomes more developed and in widespread use throughout the world, broadcasters will need to consider its possible impact on the industry. James H. Cook, Jr. provides an in-depth look at this technology in their *Satellite Earth Stations and Systems* chapter.

In the *Low Power Television* chapter, Robert M. Unetich discusses low power television/translators operating on conventional TV channels and multichannel multipoint distribution service/instructional television fixed service operating in the 2500–2690 MHz band.

In his *Cable Television Systems* chapter, Walt Cicora provides a historial perspective on cable television and offers readers an opportunity to get a better understanding of the technology that supplies video to households.

6.1
MICROWAVE & STL SYSTEMS

ERNEST M. HICKIN
MYTON ASSOCIATES INC., WINCHESTER, MA
JAMES H. ROONEY III
MICROWAVE RADIO COMMUNICATIONS, CHELMSFORD, MA

INTRODUCTION

Microwave transmission is an integral part of the broadcast engineer's operations responsibilities. The studio-to-transmitter link (STL) using frequency modulation (FM) has been the backbone of the industry for decades. Equipment design allows for essentially transparent, full color analog video transmission, as well as subcarriers capable of full FM stereo audio bandwidth and multiple voice and remote control functions. Similar technology is used for satellite earth stations (TVRO) which require backhaul links, for electronic news gathering (ENG), field production and regional news networks.

The major change new to this, the *9th Edition of the NAB Engineering Handbook*, has been the advent of digital video transmission, which requires an understanding of the advantages and the problems, associated with a variety of new modulation methods.

In most cases, microwave equipment suppliers are prepared to support broadcast engineers in their design needs. For path profiling, surveying and the prediction of path performance the engineer can undertake much of the work, or these services are available from the suppliers or from consultants. The availability of contour data for much of the USA now permits much of the initial study to be simulated by computer programs. The following discussions are presented to aid broadcast engineers in their own understanding of this technology.

PATH PERFORMANCE

Free space path loss for any electromagnetic wave arises from the spreading of the wavefront radiating from the source, like ripples on a pond. The loss increases by 6 dB (4:1 in power) every time the range doubles once outside the near field of the antenna. Since a 1:2 increase in range leads to a 2^2:1 reduction in power this relationship is generally referred to as the *inverse square law.*

Free space path loss (L), which is independent of ground or atmospheric effects, is given by:

$$L(\text{dB}) = 36.6 + 20 \log F(\text{MHz}) + 20 \log D(\text{miles})$$

or

$$L(\text{dB}) = 32.45 + 20 \log F(\text{MHz}) + 20 \log D(\text{km})$$

This is the loss between isotropic antennas (theoretical antennas which radiate or receive equally in all directions). The gain (G) of a microwave antenna is then expressed in dBi (gain relative to an isotropic antenna). For a parabolic reflector antenna, diameter d (ft), with an efficiency of 55% (which is typical of all but the smallest antennas), gain is given by:

$$G(\text{dBi}) = 20 \log d(\text{ft}) \times 20 \log F(\text{MHz}) - 52.6$$

Strictly speaking, an antenna being a passive device cannot possess power gain. The value G is the amount by which the radiation in a desired direction has been increased by redirecting energy, which would have been radiated in unwanted directions by an isotropic antenna. It should be noted that below 1 GHz it is usual to express gain relative to a dipole where 0 dBd = 2.2 dBi; thus the gain of a VHF or UHF TV transmitting antenna is most likely to be quoted in dBd.

The ratio of received power to transmitted power between two correctly aligned antennas of gains $G1$ and $G2$, when D miles apart, will be given by:

$$\frac{\text{Pr}}{\text{Pt}} = +G1 - 36.6 - 20 \log F - 20 \log D + G2 \text{ (units of dB)}$$

This number will be negative; the numerical value (i.e., absolute value) is referred to as *path loss.*

As examples, if the antennas are 6 ft diameter and the path is 30 miles long then:

at 2 GHz:

$$\frac{\text{Pr}}{\text{Pt}} = +28.5 - 36.6 - 66.0 - 29.5 + 28.5$$
$$= -75.1 \text{ or path loss} = 75.1 \text{ dB}$$

at 7 GHz:

$$\frac{\text{Pr}}{\text{Pt}} = +39.8 - 36.6 - 76.9 - 29.5 + 39.8$$
$$= -63.4 \text{ or path loss} = 63.4 \text{ dB}$$

Similarly, if the antennas are 2 ft diameter and the path is 8 miles long then:

at 23 GHz:

$$\frac{\text{Pr}}{\text{Pt}} = +40.7 - 36.6 - 87.2 - 18.1 + 40.7$$
$$= -60.5 \text{ or path loss} = 60.5 \text{ dB}$$

Note that in these equations doubling the frequency increases the total antenna gain by 12 dB (2 × 6) while the path loss increases by only 6 dB. This means that lower frequency systems require larger antennas or more power, or both, for a given received carrier level.

Other Sources of Path Loss

The free space loss is easily calculated and antenna manufacturers issue slide rules from which that loss and the antenna gain, can be read. However, atmospheric and other conditions will cause the loss to vary with time and these effects must be considered when planning a link.

Multipath (wanted signal arriving by other than the direct path) can be caused by reflections from water, hills, buildings or atmospheric discontinuities. Depending on the time delay these can add to or cancel the direct path signal, causing increases of up to 6 dB or reductions of more than 50 dB. The longer the delay, for a given magnitude, the more serious the effect as the path loss becomes increasingly frequency selective. By altering the ratio of the carrier to the sidebands in an FM system the effective deviation and therefore the level of the demodulated signal, can be raised. In a digital system, the delayed signals introduce intersymbol interference. Thus, multipath will introduce distortion into either system.

Delayed signals can also be generated within the system by reflections from discontinuities at branching devices, flanges in the feeder, at the antenna connector or by the antenna itself; these sources are very important, especially in the case of digital systems.

Atmospheric Bending of the signal due to abnormal changes in temperature and/or humidity with height can cause loss of signal due to diversion from the desired direction. This effect is discussed under the *K factor*.

Rainfall and to a lesser extent snow and ice, can attenuate the signal; this effect increases with frequency and is the over-riding factor at 18 GHz and above.

To offset these effects a system is designed to give a more than adequate signal level at the receiver under the free space loss condition (which is the norm to be expected for at least 90% of the time). The excess of signal over the minimum required for satisfactory service is called the ***fade margin***. Typical fade margins are in the range 26 to 46 dB and ideally will be larger for higher frequencies, longer paths and over water or similar difficult situations.

The choice of fade margin may involve compromises, for example when the need to use existing towers or masts may limit the size of antenna. It should also be influenced by the environment, being more necessary in humid, flat country and less necessary in dry, mountainous regions.

Outages: Multipath

Outage time is the time out of service due to propagation effects—it is usually given in minutes per month or per year. Availability, expressed as a percentage of a year or a month, is (525,960 − outage time) × 100 when the outage time is given as minutes/year, or (43,830 − outage time) × 100 when given in minutes/month. Both expressions are in frequent use.

An estimated time can be derived from an accepted formula which assumes that the path has adequate clearance for the area and is based on frequency, path length, fade margin and two empirical factors related to terrain roughness (the rougher the terrain the better as this tends to keep the atmosphere unstable) and to humidity (where the lower the humidity the better as moisture enhances atmospheric effects).

The formula, derived by Barnett & Vigants from the average of many Bell System paths, is:

$$t = a \times b \times 2.5 \times 10^{-6} \times f \times D^3 \times 10^{(-10/F)}$$

where

- t = time out of service as a fraction; in this case availability = $(1 - t) \times 100\%$
- a = {4 for very smooth terrain; 1 for average terrain with some roughness; 0.25 for mountainous, very rough or very dry areas}
- b = {0.5 for Gulf coast or similar low-lying, humid areas; 0.25 for normal interior, temperate or northern areas 0.125 for mountainous or very dry areas.}
- f = frequency in GHz
- D = path length in miles
- F = fade margin in dB

This formula can be useful in predicting performance for links where D is between 10 and 40 miles and for fade margins between 25 and 55 dB.

Outages: Rainfall

Attenuation due to rainfall is not covered by the B & V formula, which was derived from 4 GHz data where rain effects are negligible. Rainfall becomes of increasing importance at frequencies above 10–12 GHz and in higher rainfall areas. Note that it is the peak rainfall rather than the annual average, expressed in inches or mm/hour, which is important; while a day-long drizzle and a short thundershower may deposit the same volume of water it is the latter which will cause an outage.

To calculate probable outages due to rainfall a five-step procedure is used.

1. The fade margin is divided by the path length in kilometers (miles × 1.6).
2. By reference to Figure 6.1-1, from ITU-R (formerly CCIR) Report 721-2, the average rainfall rate (in mm/hr) along the whole path which would just cause a loss equal to the fade margin can be read by laying a rule from the right-hand scales through the center scale to the left-hand scale (note that this number is different for vertical and horizontal polarizations).

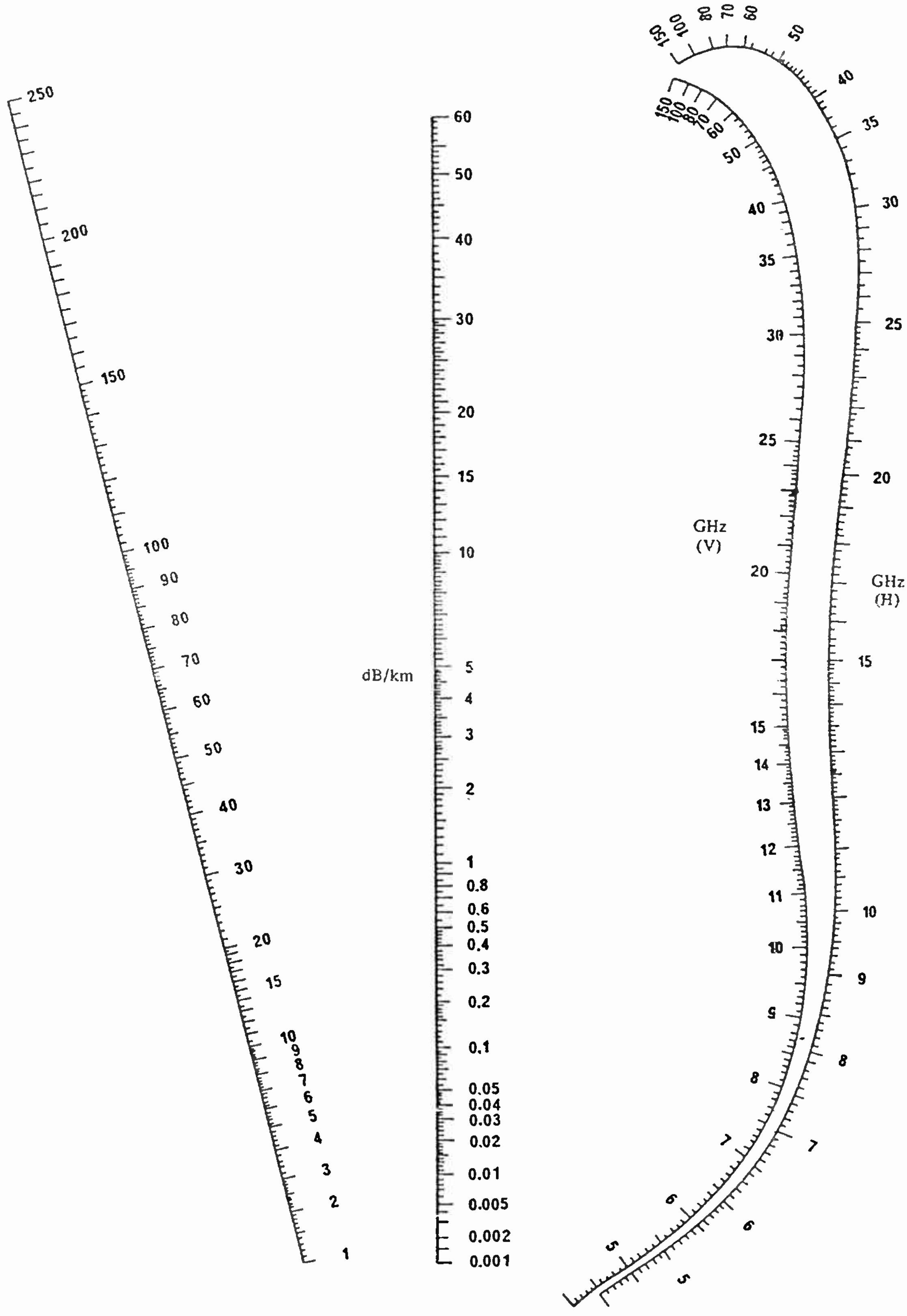


Figure 6.1-1. Specific attenuation due to rain. (H): horizontal polarization, (V): vertical polarization. *(From CCIR Report 721-2)*

3. By reference to Figure 6.1-2, from Barsis & Samson, find the point corresponding to that average and to the path length in question. Read the ratio number at the left and use it to divide the average rainfall rate to get the point rainfall.
4. By reference to Figure 6.1-3 note the rainfall zone in which the link is to be established.
5. Finally, read the outage time by reference to Table 6.1-1.

For longer paths at 13 or 15 GHz and for all paths at 18 GHz and above, the rainfall outage predominates and multipath effects as derived from the B & V formula can usually be ignored.

K Factor

The pressure and hence the density of the atmosphere surrounding the earth varies with height, getting less as the height increases and the weight of the air above decreases.

As a result the dielectric constant also decreases with height and this has a prismatic effect causing microwaves (and light waves) to bend towards the earth. Under normal conditions, the bending is less than the curvature of the earth but nonetheless micro-

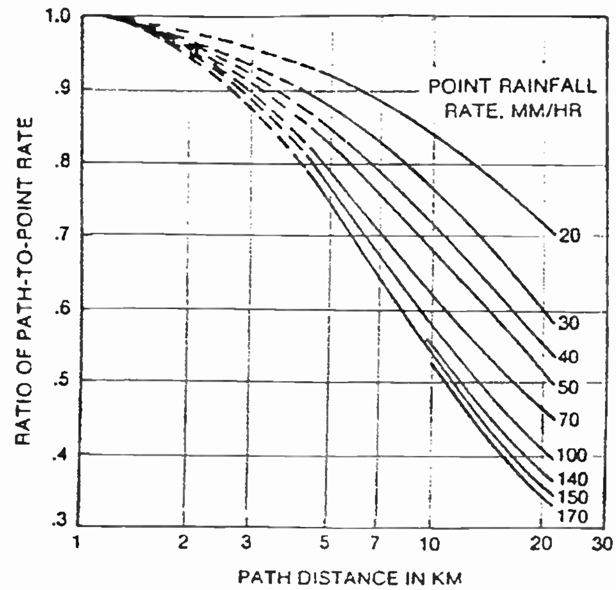


Figure 6.1-2. Path average-to-surface point rainfall rate as a function of path length. *(From Barsis & Samson, IEEE Transactions on Communications, April, 1976)*

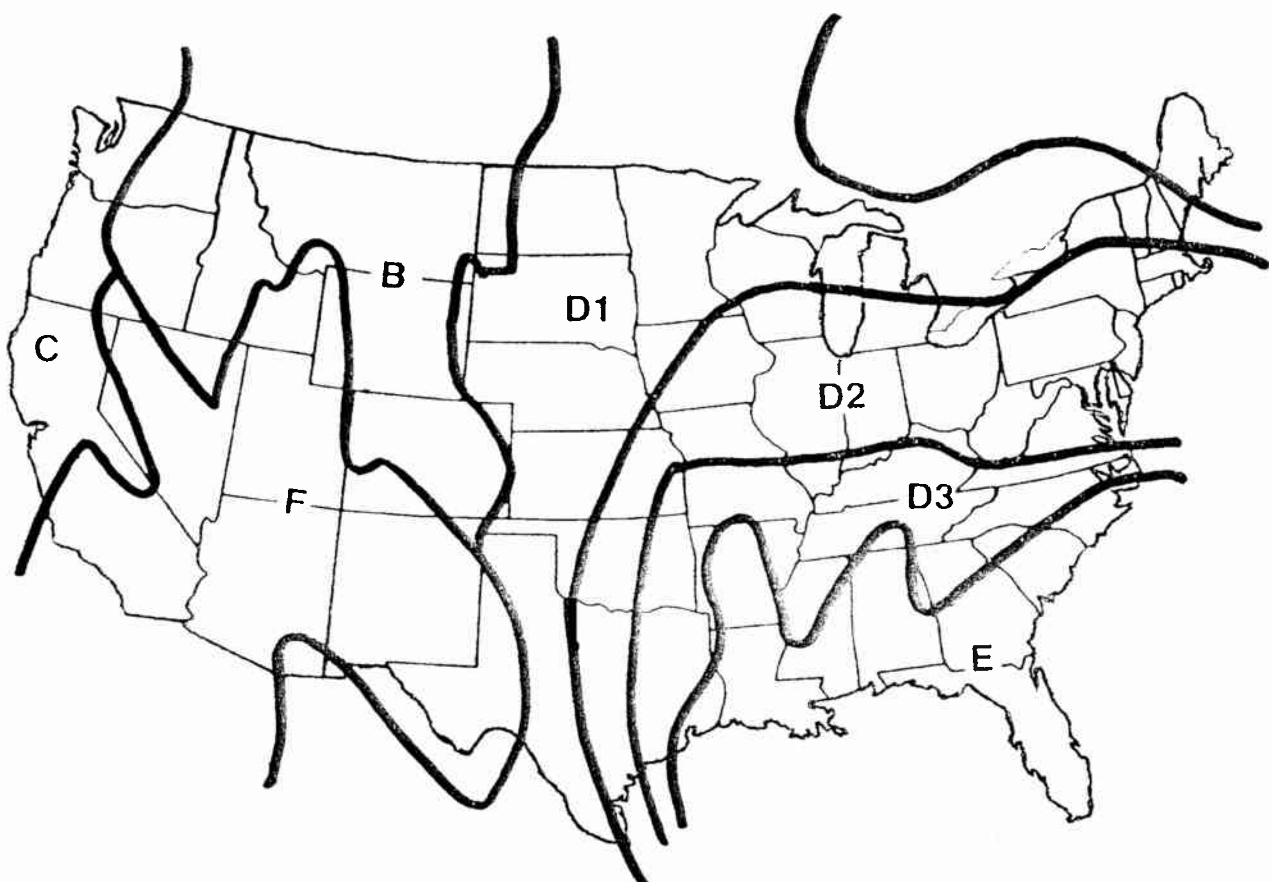


Figure 6.1-3. Geographic regions with similar rainfall statistics. *(From Crane & CCIR)*

Table 6.1-1
Point rain rate distribution values (mm/hr) as a function of percentage of a year that a given rain rate is exceeded.

PERCENT (YR)	RAIN CLIMATE REGION										MINUTES (YR)	HOURS (YR)
	A	B	C	D_1	D_2	D_3	E	F	G	H		
0.001	28.0	54.0	80.0	90.0	102.0	127.0	164.0	66.0	129.0	251.0	5.3	0.09
0.002	24.0	40.0	62.0	72.0	86.0	107.0	144.0	51.0	109.0	220.0	10.5	0.18
0.005	19.0	26.0	41.0	50.0	64.0	81.0	117.0	34.0	85.0	178.0	26.0	0.44
0.01	15.0	19.0	28.0	37.0	49.0	63.0	98.0	23.0	67.0	147.0	53.0	0.88
0.02	12.0	14.0	18.0	27.0	35.0	48.0	77.0	14.0	51.0	115.0	105.0	1.75
0.05	8.0	9.5	11.0	16.0	22.0	31.0	52.0	8.0	33.0	77.0	263.0	4.38
0.1	6.5	6.8	72.0	11.0	15.0	22.0	35.0	5.5	22.0	51.0	526.0	8.77
0.2	4.0	4.8	4.8	7.5	9.5	14.0	21.0	3.8	14.0	31.0	1052.0	17.50
0.5	2.5	2.7	2.8	4.0	5.2	7.0	8.5	2.4	7.0	13.0	2630.0	43.80
1.0	1.7	1.8	1.9	2.2	3.0	4.0	4.0	1.7	3.7	6.4	5260.0	87.66
2.0	1.1	1.2	1.2	1.3	1.8	2.5	2.0	1.1	1.6	2.8	10520.0	175.30

(From Crane, IEEE Transactions on Communications, Sept, 1980)

waves will go further than simple geometry would suggest. A convenient way to allow for this when drawing profiles is to increase the radius of the earth until the microwaves appear to be travelling in straight lines.

The ratio of this effective earth's radius to the true earth's radius is called k and its value is approximately 4/3 or 1.33 for over 90% of the time in most parts of the world. However, there are times when "k" can be anything from infinity to as low as 0.45. When "k" is equal to infinity, the earth appears to be flat; it is a condition where mirages are seen and radar echoes are received from hundreds of miles away. It is an embarrassment as interference between systems is increased—fortunately, it is relatively rare. Values of "k" between 1 and 0.45 can occur for a few percent of the hours in a year and it is necessary to allow for this if a reliable link is to be established. Figure 6.1-4 is a map of the continental United States showing contours of equal minimum k-factor; this is based on refractive index measurements made by the Central Radio Propagation Laboratory.

Clearance Requirements

Figure 6.1-5 shows the manner in which the microwave signal is attenuated when the path is close to an obstruction. Clearance is stated as a fraction of the *first Fresnel zone (FZC)*, a function of frequency.

FZC (feet) is given by:

$$\mathrm{FZC} = 2280(d(D - d)/fD)^{1/2}$$

Where

f = frequency (MHz)
D = path length (miles)
d = distance to the obstruction from either end of the path.

Figure 6.1-5 illustrates that the attenuation with a clearance of 0.6 FZC is equal to the free space attenuation. However, as noted earlier, k-factor variations will mean that more than this clearance will have to be built in to allow for values less than 4/3 and a typical design parameter is to plan for 0.3 FZC for the lowest value of "k" expected on the path, as derived from Figure 6.1-4. While such a clearance will introduce 2 to 8 dB of loss, this is well within the fade margin of a well designed link.

Path Profiles

Sites for microwave link terminals are likely to be studios, earth stations or TV transmitters, all of which are predetermined. Choice of sites for repeaters, if they are necessary, may be constrained by availability and access. Since the access road is often the most costly part of a site, it can prove more economic to avoid the highest point and use a taller mast close to a road.

To see if any combination of sites is viable and to determine the antenna heights required at each end of the path, it will be necessary to construct a profile. This can be done with the aid of USGS 7.5 or 7.5 × 15 minute topographical maps. Now that the entire country has been surveyed to 3 arc-seconds and this

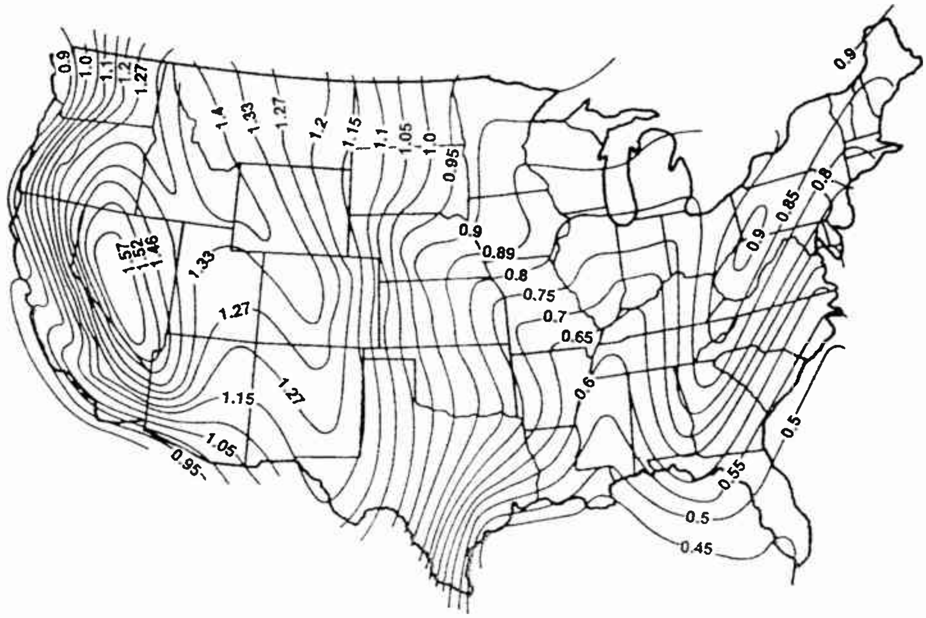


Figure 6.1-4. Estimated minimum K-factor for the continental United States.

topographical data is available on disc, there are programs which will draw the profile given the coordinates of the terminals. The programs are expensive but consultants (and some manufacturers) have them and the work can be subcontracted economically. Whichever route is chosen, some on-site examination of the path is almost always necessary to ensure that tree heights are correct, and that there have been no new buildings added on the path since the maps were compiled or the data entered.

If the work is to be done in house the profile can be drawn on simple squared paper or on special *k*-

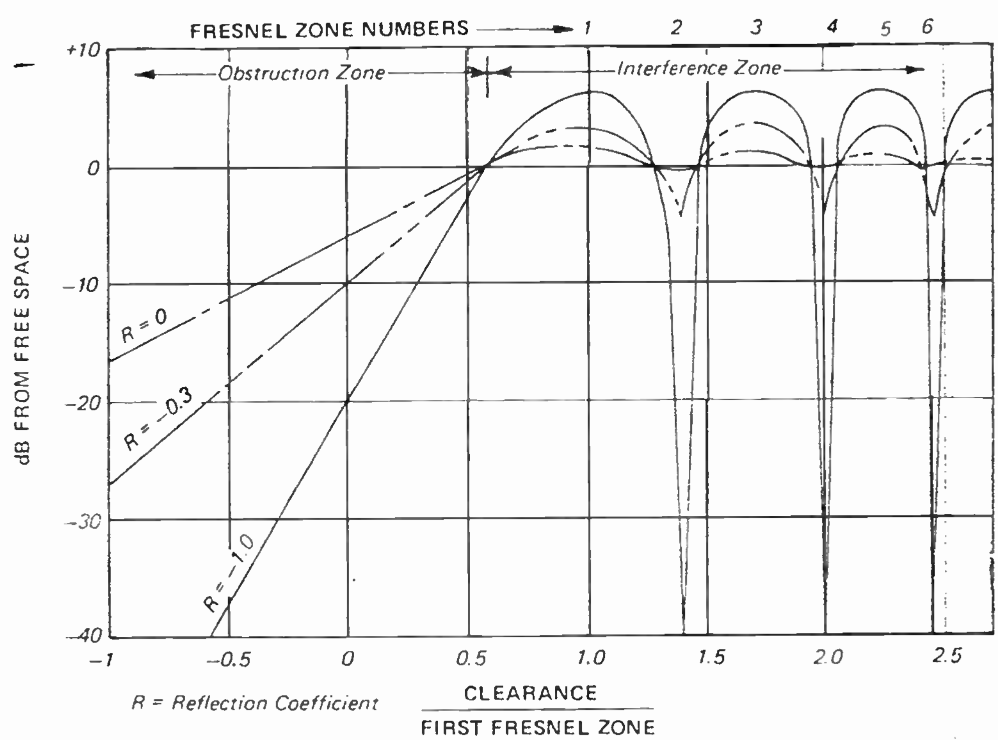


Figure 6.1-5. Attenuation vs path clearance for various types of obstruction.

factor paper (usually prepared for $k = 4/3$). Squared paper has the double advantage of easy availability and that any convenient scale can be chosen for the x and y axes; k-factor paper has the advantage that potential obstructions are more obvious and the effects of path length on necessary clearance stands out. Whichever paper is used the first step is to draw a straight line on the map between the sites. Then plot the contour heights on the paper every half mile where the ground is low, more often near high spots. It is usually clear where the dominant obstruction is going to be. Since the topographical map shows only ground level, the height of trees or buildings must be added in the region of any obstruction; this requires a site visit unless the terrain is well known.

Having established the level of the obstructions (ground level plus any trees or buildings) on the path an allowance (additional clearance) must be added to take care of the minimum clearance (0.6 FZC is recommended) and lower than normal values of the k-factor, using the following formula:

If using squared paper then add to each obstruction:

0.6 FZC + $d(D - d)/2$ (feet) (minimum clearance factor)

0.3 FZC + $d(D - d)\ 3k/2$ (feet) (low k-factor value)

If using 4/3 paper then add to each obstruction:

0.6 FZC (feet) (minimum clearance factor)

0.3 FZC + $d(D - d)\ 3k/4$ (feet) (low k-factor value)

where D is path length in miles, d is the distance of the obstruction from either end of the path and "k" is taken from Figure 6.1-4.

Then a straight line drawn through the highest points (usually (2) except on short paths or at lower frequencies) will pass through the minimum antenna heights at the terminals. If there are problems in mounting the antenna at one end at the height indicated then the line can be adjusted to the reduced height at that end, at the expense of increasing it at the other end. If most obstructions are near one end of the path it may be more economic to raise the antenna at that end and save a greater amount at the far end.

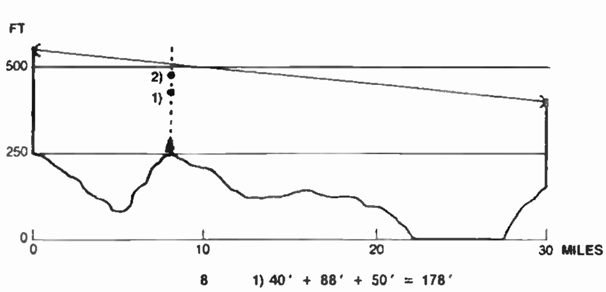


Figure 6.1-6. Example of profile plotted on squared paper, with required clearances for (1) 0.6 first Fresnel zone when K = 4/3, (2) 0.3 first Fresnel zone when K = 2/3 (50-ft tree allowance)

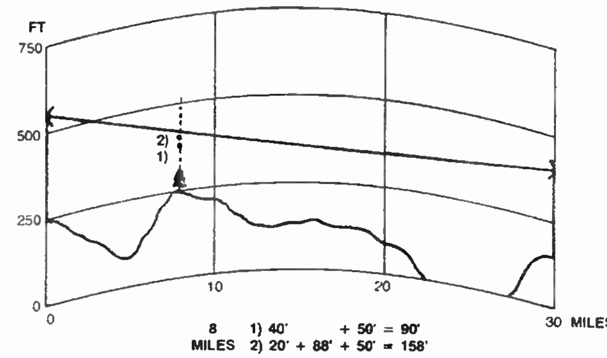


Figure 6.1-7. The profile of Fig. 6.1-9 plotted on k = 4/3 paper; (1) and (2) marked as on that figure.

Figures 6.1-6 and 6.1-7 show examples of the same profile plotted on squared and 4/3 papers and the final selection of antenna heights. The worst k-factor assumed was 2/3 shown as (2) on each profile at the main obstruction point at 8 miles.

ANTENNA SYSTEMS

Several different types of antenna can be used in a microwave TV link system. Omnidirectional antennas with vertical or circular polarization are frequently used for portable systems on moving vehicles and helicopters. Dipoles or discs, often in fiberglass covers, are used at 2 GHz for ENG either singly or in pairs; polarization can be switched to horizontal, vertical, right or left circular polarization to reduce path reflections. For STL's, intercity and community receiver antenna systems (CARS), the parabolic reflector type antenna is by far the most common. (See EIA/TIA Standard 195A for mechanical and electrical specifications.) A transmission calculation worksheet is given in Figure 6.1-8.

Isotropic

As previously noted the isotropic antenna is a hypothetical antenna used as a reference against which the gain of practical microwave antennae can be measured. Gain is the given in dBi, gain relative to an isotropic antenna. The isotropic antenna has a radiation pattern which is a perfect sphere. Other antennas are designed to enhance their radiation in a given direction at the expense of all other directions; this has the double advantage of increased sensitivity in the desired direction and reduced pickup of unwanted signals from other directions.

Parabolic

The parabolic reflector antenna comprises a feed at the focus of a parabolic reflector; the focus may be in the aperture plane of the reflector, for minimum sidelobes, or more usually in front of the aperture, to give maximum gain. The feed may be a dipole with a subreflector at frequencies up to 2.5 GHz; it can be fed by foam filled coaxial cable to avoid the need

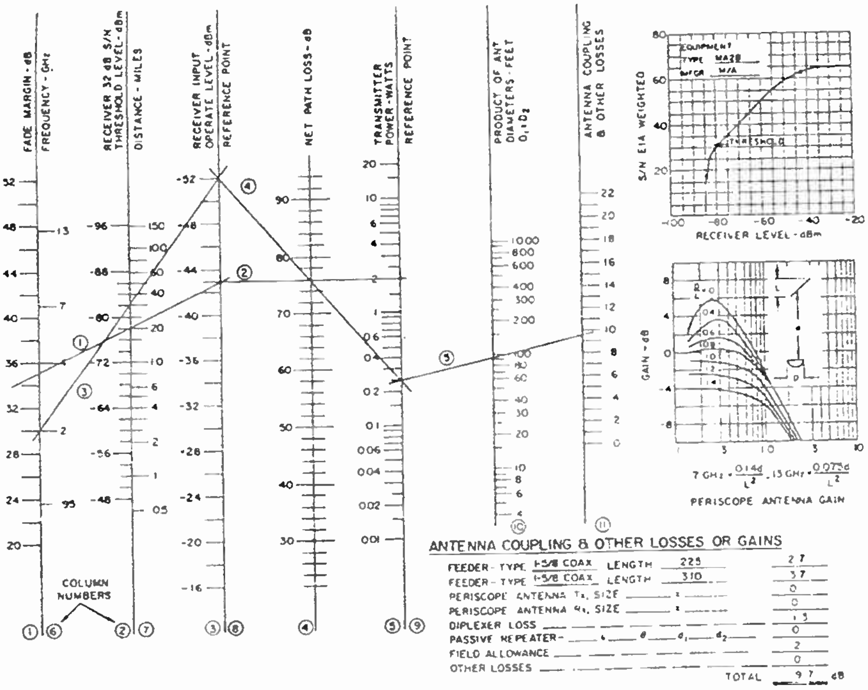


Figure 6.1-8. Transmission calculation worksheet.

for dry air pressurization. At higher frequencies, it is invariably a small horn fed by waveguide. The feed is designed to concentrate radiation onto the reflector with as little spillover as possible; a compromise has to be made between maximum gain and minimum spillover depending on the application. When minimum sidelobe radiation is critical (high performance antennas), a cylindrical metal shroud is fitted to the reflector extending beyond the focal plane; this may have absorbing material attached to the inner surface.

At frequencies up to 2.5 GHz the parabolic reflector of an antenna can be built up from closely spaced rods which must all be parallel to the polarization of the feed; this construction gives some relief from wind load. Above 2.5 GHz the reflector is solid but may be metal, metalized fiberglass or other reflecting material. Dual cross polarized feeds, with isolation in the 25 to 33 dB range, can be used to increase the number of signals sharing the antenna. Dual band antennas (for example, 7 GHz and 13 GHz) can also be built, each band having one polarization. Since the mast or tower is a costly item, these types offer reduced wind loading for multichannel systems.

The gain of a parabolic reflector antenna is given by:

$$G(\text{dBi}) = 10 \log 5.5(Df)^2$$

and the half power beam width (the total included angle between the points at which the gain has fallen by 3 dB) is given by:

$$\text{BW (degrees)} = 68.7/(Df)$$

where D = parabolic reflector diameter (ft) and f = frequency (GHz). Note that the half power beamwidth is a guide to the torsional rigidity required of the antenna mounting (including the mast or tower); in practice any movement should not exceed one quarter the above value to ensure minimum effect.

The voltage standing wave ratio (VSWR, expressed as a ratio to 1) or return loss (given in dB), which results from mismatch between the feed and free space, varies with antenna size and design. A standard antenna of 4 ft diameter or more may have a 1.10 VSWR (26 dB) while a high performance antenna would have a 1.06 (30.7 dB) to 1.04 (34.2 dB) figure.

Digital radio links, being more sensitive to reflections in the transmission path, are likely to require high performance or low VSWR antennas.

Radomes

Radomes can be fitted to parabolic antennas, the design depending upon the antenna construction. In addition to giving weather protection, the radome tends to reduce wind loading and torsional instability caused by the aerofoil effect of the reflector surface. For standard antennas, the radome is usually convex in shape and made of fiberglass, which may have imbedded heating elements. High performance antennas with shrouds have plane covers of Teglar or Hypalon stretched across the aperture; because these flex in windy conditions they tend to shed ice deposits and are not heated.

Other Radiation Patterns

Fixed point-to-point links, for STL applications as an example, use directional antennas to reduce power requirements, unwanted reflected signals and interference to and from other users of the frequency band. However, for portable links as used for ENG and other field applications wide-angle antennas may be needed. For links from motorcycles or cars to helicopters, hemispherical radiation patterns are used. Hand held cameras can have antennas with omnidirectional radiation in the horizontal plane. Manufacturers of portable equipment provide special antennas for differing applications.

Antenna Feeders

Microwave equipment is connected to the antenna with coaxial cable or waveguide. Cable is used below about 3 GHz and may be foam filled or with air dielectric and minimal plastic support. The copper outer is typically corrugated with a protective outer coating. Waveguide can be rigid, with rectangular or circular cross section, but as with coaxial cable is generally made of corrugated elliptical copper with a protective outer coating. The corrugations both strengthen the feeder and make it bendable. However, both cable and waveguide can be kinked if not installed with care and bending jigs are strongly recommended. Both Andrew Corporation and Cablewave Systems provide a wealth of information on feeders, connectors and all the accessories needed for installation in their catalogs.

When waveguide or air-dielectric cable is being used, it must be maintained at a positive pressure (0.5–5.0 lb/in^2) above atmospheric with dry air or nitrogen. Compressor/dehydrator units to supply this dry air can be fully automatic (where the desiccant is dried out periodically by the unit) or semi-automatic (where the desiccant must be periodically dried out or replaced by an operator). To avoid rapid cycling of the compressor the volume of air in the system should not be less than 1.5 cubic feet: this can be achieved in small systems by fitting a regulating tank.

Connectors for both cable and waveguide need careful fitting to avoid leakage of air and ingress of moisture. The mismatch introduced by a connector can be reduced by the use of tuned connectors, which have provision for canceling reflections.

Passive Repeaters

When two terminals are less than 10–15 miles apart, but the direct path is blocked by an obstruction, it may be possible to employ a passive repeater (a microwave mirror) to bypass the obstruction. The passive design is dependent upon the included angle between to two sections of the path at the repeater site. If the angle is less than about 120° a single flat panel can be used as the reflector—typical sizes at 7 or 13 GHz are from 8′ × 10′ to 20′ × 32′. Even larger ones can be used at 7 GHz while special smaller ones with a carefully controlled surface are available for 18 and 23 GHz.

For angles greater than about 120°, an arrangement of two closely spaced flat reflectors, or two parabolic antennas back-to-back with a short section of wave guide between them, can be used. In the latter case, a low noise amplifier can be fitted in the waveguide to improve performance. Care must be taken to prevent overloading of the amplifier if the repeater is near one terminal of the system.

Figure 6.1-9 is a nomogram with which the transmission loss through a passive repeater (relative to the direct free space loss) can be read. It will be clear that the loss is significantly less at higher frequencies and passive repeaters are seldom used below 4 GHz. Because the passive repeater requires no power, and little maintenance, its use should not be overlooked when designing microwave systems. (For more information

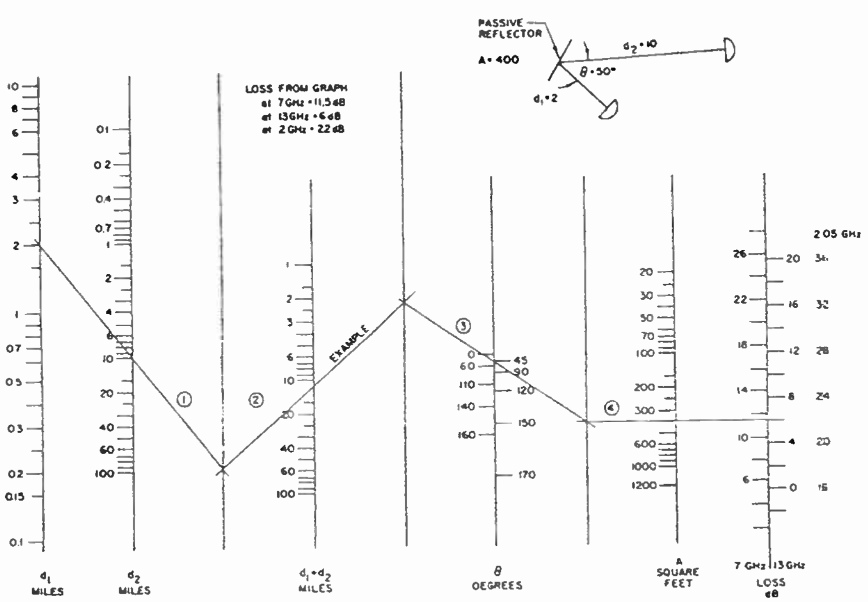


Figure 6.1-9. Passive repeater loss relative to direct transmission distance $D_1 + D_2$.

see *Passive Repeater Engineering* by the Microflect Company.)

RADIO EQUIPMENT

There are two generic terrestrial line-of-sight microwave transmission techniques in use today, analog systems typically using FM and digital systems typically using phase-shift keying (PSK) and quadrature analog modulation (QAM, typical 16-QAM to 512-QAM). Historically, transmission of video and audio signals over the last 50 years has been almost exclusively carried out using frequency modulation of the microwave carrier by an analog signal (the exact waveform to be broadcast rather than a sampled and coded version of that waveform from which the original can be regenerated at the end terminal as is the case in digital systems).

Since the television transmitter itself has used that analog signal, it has been a logical arrangement. Now that the studio equipment increasingly generates a digital output and the new standards for digital television are firmly set on a digital format, a new look at link design is required. Audio information also has usually been frequency modulated onto subcarriers, above the video, although digital modulation of the sync pulse is widely used in Europe and digital modulation of subcarriers is increasing. With DTV, the audio is built into the digital data stream.

The microwave transmitter and receiver may still comprise similar units, but whereas there is but one set of FM parameters set out in ITU-R recommendations (making interconnection of different manufacturers equipment simple, even at RF), common digital parameters have yet to be accepted. Because the bulk of equipment in service is still using FM, these systems are covered in detail in the next section. The section which follows deals with the various modulation techniques, (analog and digital).

Analog FM Systems

Two types of transmitter are in common use, *directly modulated* RF and *heterodyne*. In the former the analog signal is applied directly to the RF source to vary its frequency; in the latter the analog signal is used to frequency modulate an intermediate frequency (invariably 70 MHz) which is then transposed to the final radio frequency.

The advantage of the directly modulated RF system is simplicity. Since the modulation is applied to the RF source, no RF amplification is needed whether that source is an oscillator at the final frequency or a lower frequency source followed by one or more multipliers. Also, in multihop systems, since the carrier is generated afresh at each repeater, there is no build up of carrier-frequency error. The disadvantage only shows up in multihop systems where multiple remodulations can lead to level instabilities and bounce; after about three repeaters it is desirable to fit a level clamp.

The advantages of the heterodyne system lie in the multihop case. Once the modulation process has been accomplished at the initial transmitter the deviation (and hence the baseband level after demodulation) will not change; other sources of noise and distortion associated with the modulator occur only once in a system of any length. The disadvantages are that to ensure that a carrier frequency error does not build up over a multihop system, very stable oscillators, or an AFC system, are required. As the transmitter involves a mixing process, the power level is low and considerable RF amplification is needed. These factors lead to higher costs. The heterodyne repeater is the system of choice for longer microwave systems.

Directly-Modulated Transmitters

A basic transmitter of this type is shown in Figure 6.1-10. It comprises a video amplifier (with preemphasis), an oscillator, which can be frequency, modulated, a crystal referenced AFC loop and the options of a power amplifier and/or multiplier to deliver the required power and frequency. Various other optional items are available to extend the usefulness of the transmitter; these include audio subcarrier modulators, pilot generator, hot-standby protection switch, and off-air monitor, while portable equipment may have multi-RF channel switching. In a fixed link, the output will have a channel filter and provision to combine the signal with other transmitters and/or receivers sharing the same antenna and feeder.

The baseband amplifier will be broadband (10 MHz or so) to ensure linearity and permit multiple subcarriers above the video signal. Preemphasis is used to improve the signal-to-noise ratio and reduce the effects of the dc component of the video signal. Combining circuits are used to inject the audio subcarriers as well as a pilot and the engineer's order wire.

The transistorized voltage controlled oscillator typically runs at 2 GHz where high power is readily available. The output may be sufficient to use directly—more often, a 10 dB amplifier is used. To generate frequencies in the 6–8 GHz band, the 2 GHz signal is amplified and then multiplied by 3 or 4 times (the deviation of the primary oscillator being reduced to allow for this; for 13 GHz a further doubler is used). The power amplifier comprises one or more broad band stages to provide 8 to 10 W at 2 GHz, enabling 1–2 W at 7 GHz and 0.5–1 W at 13 GHz to be generated using varactor diodes.

The phase locked AFC uses a microwave IC divider which samples the output of the 2 GHz oscillator and compares the phase of the divided signal with a crystal oscillator; this is a simple and reliable method. Portable equipment, which may have to operate on any channel in the band, may use multiple crystals or more likely a frequency synthesizer as the reference in the AFC circuit. Since other circuits are broadband, no tuning is necessary when the reference is switched to another channel.

Because both fixed and portable equipment is likely to have to operate in the vicinity of high powered TV transmitters careful, filtering of power lines and baseband inputs is necessary.

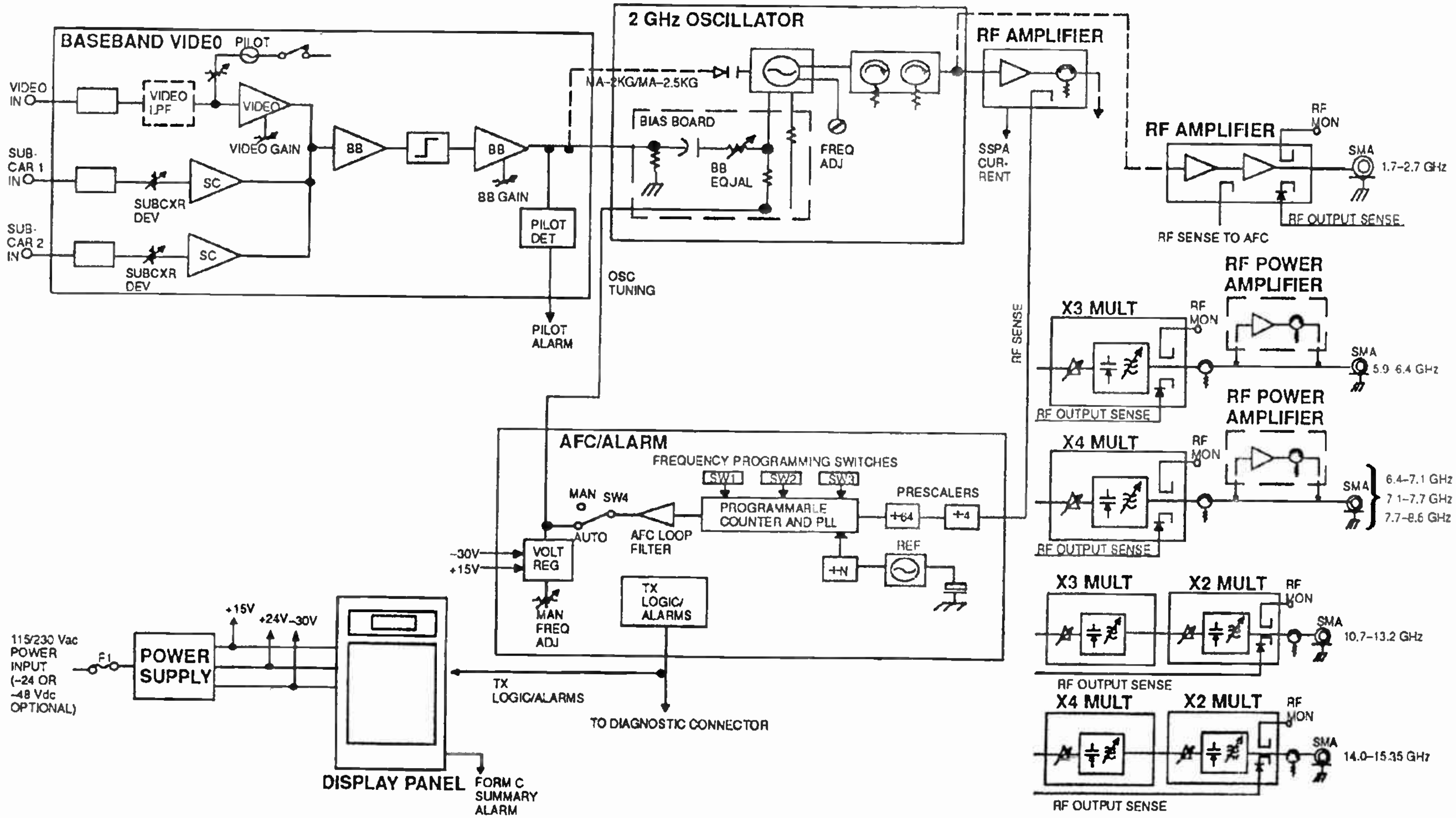


Figure 6.1-10. Directly-modulated transmitter—Functional block diagram.

Heterodyne Transmitters

A typical heterodyne transmitter block diagram is shown in Figure 6.1-11. It comprises an IF amplifier and a microwave source both of which feed an upconverting mixer. One sideband, usually the upper one, is selected by a filter or by the use of an image rejecting mixer. The mixer output, at a level around 0 dBm, is then amplified by 30 to 40 dB to provide the transmitter output. At a terminal the modulation is applied to a 70 MHz frequency modulator which produces a constant voltage at a frequency proportional to the applied voltage; the deviation sensitivity is such that a 1 V peak-to-peak signal produces a deviation of 8 MHz peak-to-peak. The modulator has a baseband amplifier with provision for the injection of subcarriers identical to those feeding the oscillator of the directly modulated transmitter. At a repeater, the 70 MHz output from the receiver is used instead of that from the modulator. To maintain transmitter output in the event of a failure of the incoming IF signal the IF amplifier is preceded by a monitor which triggers a locally generated 70 MHz carrier *(carrier reinsertion)*; this is important as the RF output is not being regenerated at each transmitter but is an amplified version of the original input to the system.

Note that in Figure 6.1-11 the local oscillator has provision for modulation; this is an important feature which permits the injection of a single sideband suppressed carrier (SSB-SC) signal with order wire or supervisory signals even though the through signal has not been demodulated. Since this follows the carrier reinsertion feature, alarms generated at a repeater will still be sent out even if the preceding receiver output has failed.

Receivers

The same basic receiver design is used with both directly modulated and heterodyne transmitters. Figure 6.1-12 is a block diagram of a typical receiver. In fixed (STL) equipment the RF input is invariably via a channel filter; this facilitates multiplexing with other receivers or transmitters, suppresses local oscillator radiation and eliminates any response at the image frequency (140 MHz above or below the received signal). In Figure 6.1-12, the input is amplified at RF in a low-noise amplifier (LNA) before the image rejecting mixer. Following the mixer is an IF amplifier, a filter equalizer which provides a flat passband with high adjacent channel rejection and a further high gain amplifier with automatic gain control (AGC). This latter feature ensures a constant level of 70 MHz at the receiver output despite level changes of 50 to 60 dB *(dynamic range)* in the received carrier level (RCL).

In the heterodyne repeater one IF output is used to drive the following transmitter via the carrier reinsert unit; another output may be used to drive a demodulator for the SSB-SC orderwire and supervisory channel or for monitoring.

The local oscillator may be a Gunn or dielectric resonance (DRO) type oscillator. In the remodulating case a stability of 0.005% is sufficient as any errors are not passed on to the following transmitter. In the heterodyne case frequency errors will add up and either a stability of 0.001% or better, or an AFC circuit which will ensure a 70 MHz output, are necessary. Figure 6.1-12 shows the latter arrangement.

Optimum receiver noise performance is achieved when the minimum necessary bandwidth is used. This is typically 20–25 MHz between the 3 dB points for TV with up to four audio subcarriers and a service channel. For portable equipment, with one or at most two audio subcarriers, channel allocations may require an IF bandwidth of 12.5 MHz. Recently, some major educational networks have been established with 4.2 MHz video bandwidth and a single subcarrier at 4.5 MHz, using a 10 MHz IF bandwidth; by reducing the peak deviation the quality can be maintained close to broadcast standards. In all cases, the characteristics of the filter equalizer must be carefully controlled for good phase and amplitude linearity.

The main IF amplifier follows the filter and provides 50–65 dB of gain. Since the input microwave level can vary between −20 and −85 dBm, while the output must remain substantially constant at about +5 dBm, an automatic gain control (AGC) must be incorporated in the main IF amplifier. When very large variations are expected, as in portable applications where path lengths can be very short, the AGC action may be extended to the LNA; this can prevent overloading of the mixer. The receiver should have multiple IF outputs available to allow for repeaters where the signal may be required to go in more than one direction as well as feeding a demodulator; for preference these outputs should be available squelched or unsquelched.

The demodulator comprises a limiter, discriminator, deemphasis and a baseband amplifier. The limiter is an essential part of an FM receiver in that it removes AM before the discriminator; threshold performance is significantly affected by limiter design. The discriminator demodulates the FM signal back to AM. In a TV system the discriminator must be linear over a wide bandwidth about the 70 MHz IF, at least ± for video and more when multiple subcarriers are being carried. Any non-linearities will produce distortion and noise by generating intermodulation products.

Finally the deemphasis network and baseband amplifier restore the baseband response and ensure the required 1 V peak-peak level of the video signal. The video output will require a lowpass filter to remove the sub-carriers; typically 35–40 dB is used.

Subcarriers

In addition to carrying the video traffic, a microwave link can carry other information above the video portion of the baseband. As the NTSC video bandwidth does not extend above 4.5 MHz, the region above this, after allowing a guard band, is available for other traffic. Four different uses are made of this region: one or more audio program subcarriers; an engineering order wire; supervisory signals (alarms and controls) and a continuity pilot. All these generally deviate the carrier at 20–26 dB below the 4 MHz video peak.

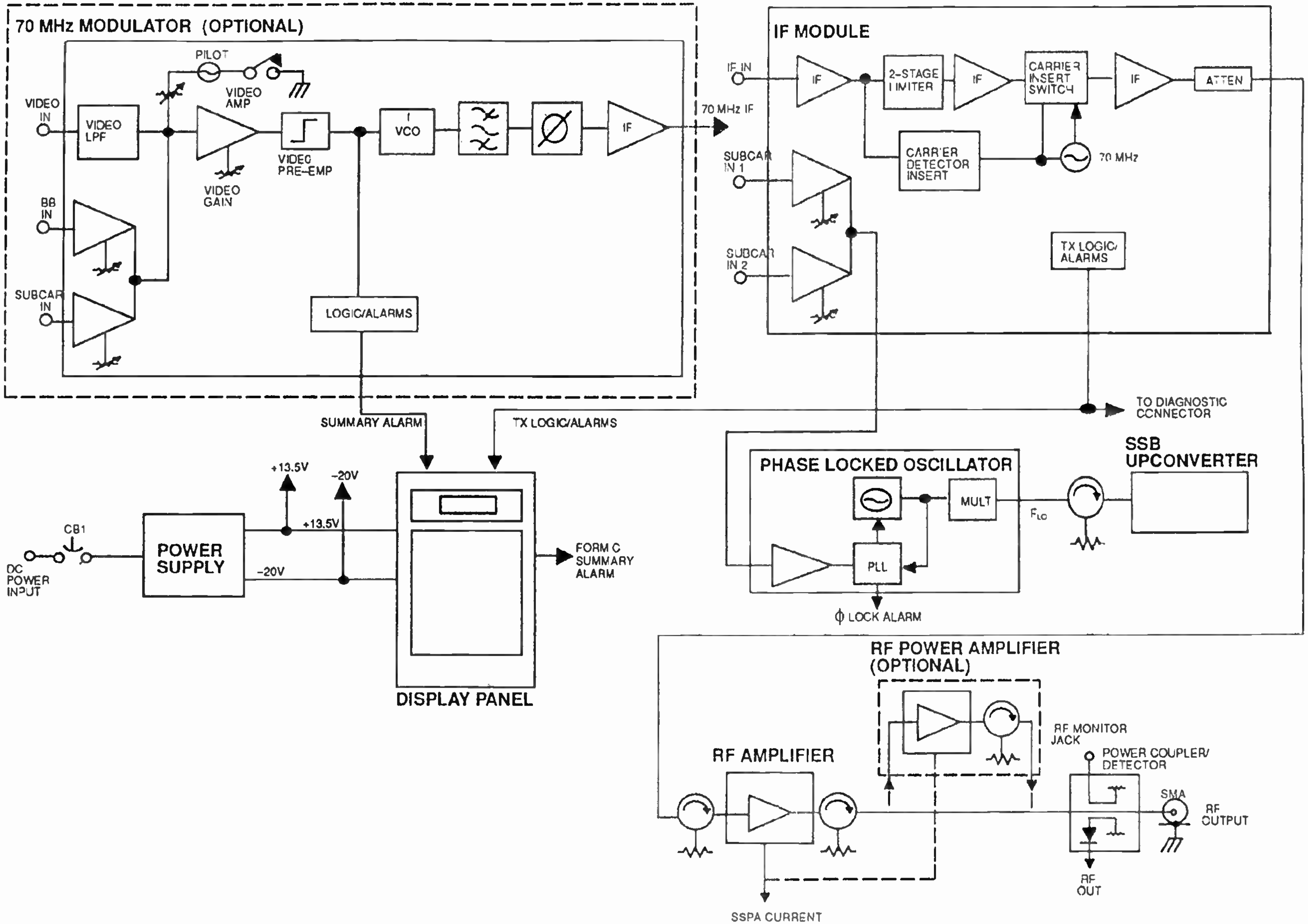


Figure 6.1-11. Heterodyne transmitter—Functional block diagram.

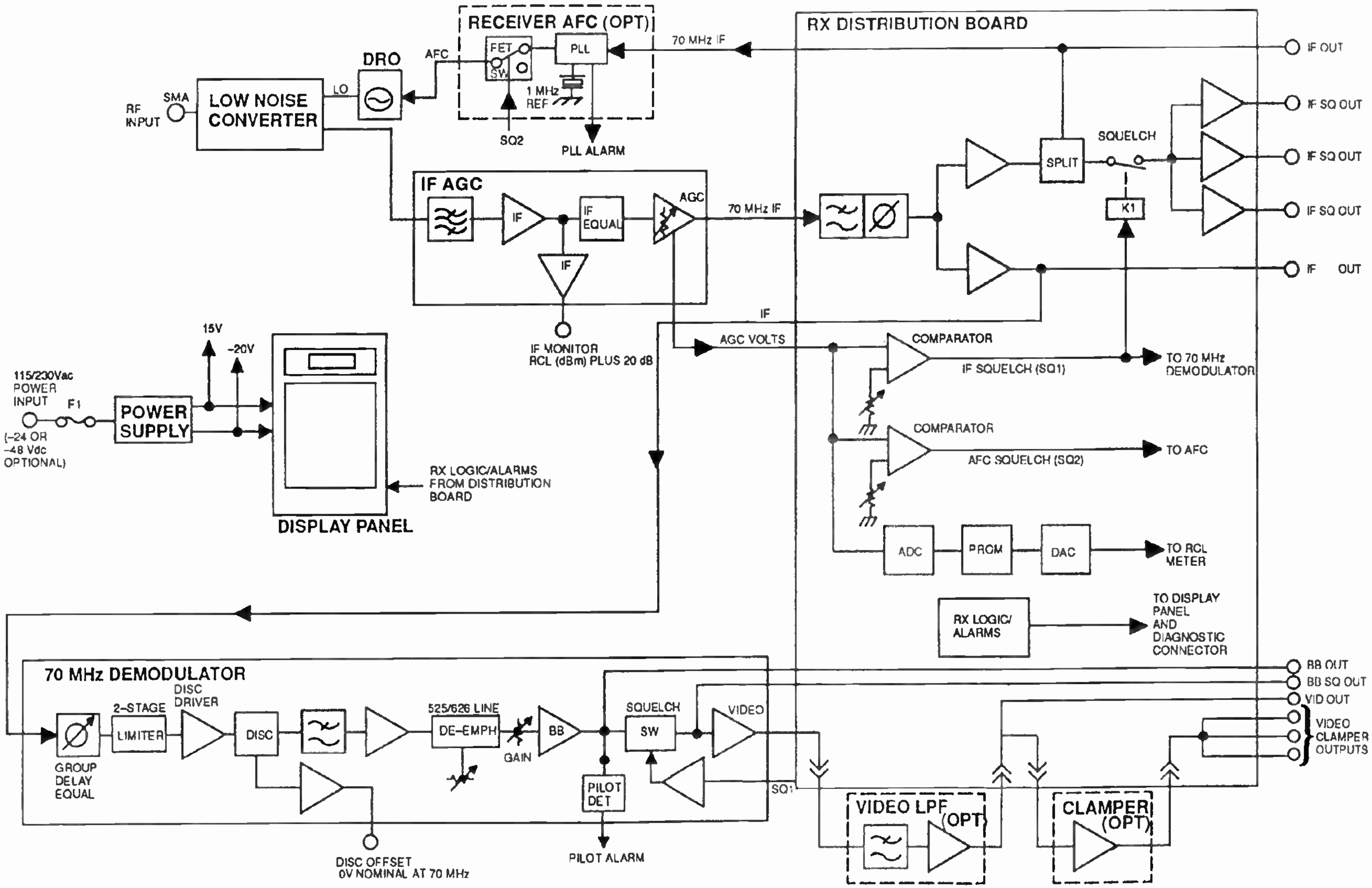


Figure 6.1-12. Receiver—Functional block diagram.

FM audio subcarriers are generally deviated at 75 kHz peak at 1 kHz; to improve the signal-to-noise ratio the audio signal is preemphasized. Maximum audio bandwidth is 15 kHz, giving an information bandwidth using Carson's rule of 180 kHz. To allow for filtering a minimum subcarrier spacing of 500 kHz is used, with increasing separation at higher frequencies to avoid beats between subcarriers falling into other channels. Typically, a 15.7 kHz trap filter is fitted to reduce the effects of intermodulation with the sync pulses.

Up to four FM subcarriers are used with NTSC STL links, at 6.2, 6.8, 7.5 and 8.3 MHz (with 4.83 and 5.2 MHz being used with ENG systems where the RF bandwidth is restricted). PAL and SECAM systems with 5.5 or 6.0 MHz video bandwidth use 7.020, 7.5, 8.065 and 8.59 MHz. These can carry stereo left and right and second language and FM transmitter feeds to co-sited transmitters. An encoded stereo signal (left plus right, left minus right, second audio channel and pro channel) can be carried on a single subcarrier, permitting the encoder to be in the studio; this requires a special modulator and demodulator.

Program audio equipment may be supplied as a separate unit or packaged in with the radio terminals. In the FM case the transmit end comprises a baseband amplifier with level setting controls to permit the recommended deviation to be achieved with inputs from 0 to + 18 dBm, a preemphasis network and a frequency modulated oscillator. The receive end comprises a filter followed by either a subcarrier amplifier or a mixer and IF amplifier, limiter/discriminator, deemphasis network and baseband amplifier with level setting capability similar to that at the transmit end. Unlike the video case there is no universally accepted audio interface level; the impedance is invariably 600 Ω balanced at the transmit end and the same, or a low impedance, at the receive end.

Often a SSB-SC AM signal is carried above the audios, at 8.59 or 8.85 MHz (for example). By using a suppressed carrier, information can be inserted at repeater stations without carrier beats. In a typical SSB-SC system the bandwidth can be up to 108 kHz, with 0.3–4.0 being used for an omnibus engineer's order wire, 4–12 kHz for supervisory and control signals and 12–108 kHz for up to 24 ITU-T (formerly CCITT) telephone channels.

In addition, a continuity pilot at the ITU-R recommended frequencies of 8.5 or 9.023 MHz is also used; this is essential if a protected system is to be provided. In the broadcast bands, only hot standby protection is permitted. In this system two transmitters on a single frequency run continuously with an RF monitor controlling a changeover switch to ensure a working transmitter is connected to the antenna. Similarly, two receivers are coupled to the antenna with a splitting network; a baseband switch controlled by the pilot ensures continuity.

Power Supplies

Microwave equipment requires voltages in the 5 to 30 VDC range. These can come from 60 Hz ac or from battery banks, usually − 24 or − 48 VDC.

Figure 6.1-13 is a block diagram of a typical power supply system. Alternate current units comprise a transformer and rectifiers to obtain the unregulated operating voltages. Various types of transistor regulator are used to keep the operating voltages within a narrow range. Direct current units, when used, are chopper types to reduce the transformer size and improve efficiency. Good designs tend to employ protection circuits on both input and outputs to guard against overloads due to lightning surges and accidental short circuits.

Digital Systems

With the advent of digital video, microwave transmission links will remain an integral part of the broadcast engineer's operation responsibilities. However, the existing FM radio infrastructure currently in service will not be capable of supporting the digital modulation needed for digital video transmission.

The STL will remain the industry backbone. During the analog to digital transition period, the STL in some cases will be required to carry a HDTV signal and a digitized NTSC signal. The simultaneous transmission

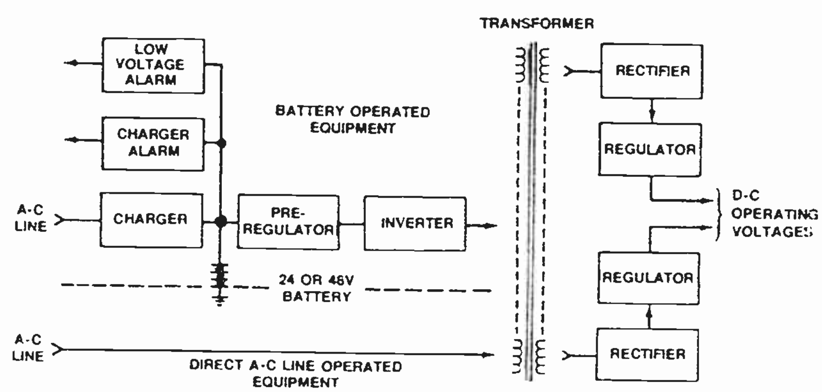


Figure 6.1-13. Power supply—Functional block diagram.

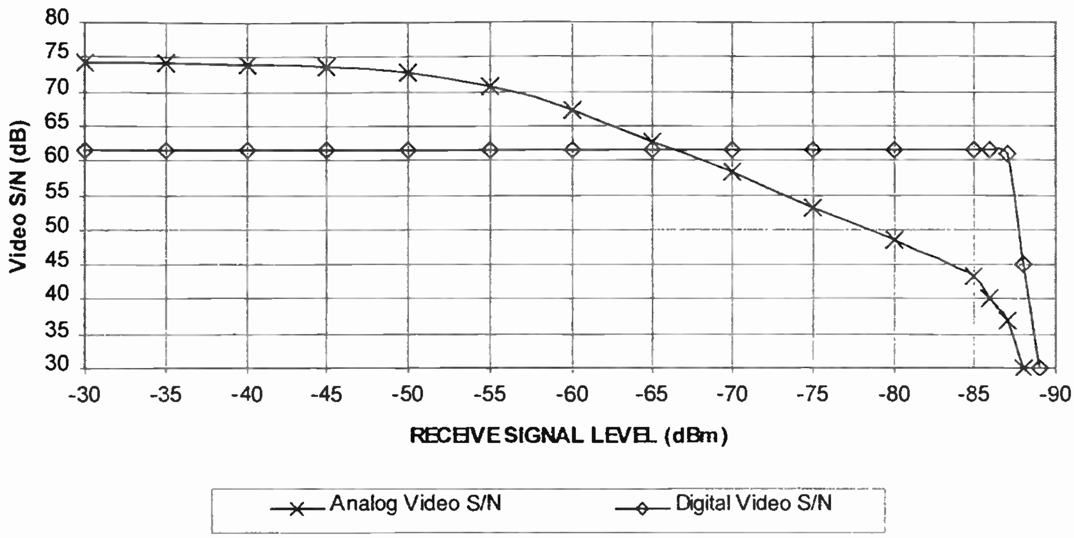


Figure 6.1-14. Typical Analog Video S/N vs. Digiral Video S/N.

of both signals is made possible by the emergence of MPEG-2 (Moving Picture Experts Group) technology and time division multiplexing.

In a properly designed digital video microwave link, the major advantage to be had over the analog link is that the video signal-to-noise (S/N) ratio is constant over the microwave receiver's operating range. This inherent digital feature provides constant picture quality over the radio receiver's dynamic range. Figure 6.1-14 shows the S/N comparison between an FM analog and 16-QAM DS3 radio system. The digital S/N is limited by the analog-to-digital (A/D) converters employed in the video encoding system and not the received signal level (as in the analog case), as illustrated by the figure. The rule of thumb is 6 dB of S/N for every bit used in the A/D. For the digital system show in Figure 6.1-14, the video encoder employs a 10 bit A/D.

In the coming years, there will be a great temptation to replace analog radios with digital radios in many of the existing links around the country. However, simply replacing an analog radio with a digital radio will not always provide the performance shown in Figure 6.1-15. It is the responsibility of the broadcast engineer to insure the path will meet acceptable availability criteria. Fortunately, the broadcast engineer will not be alone during this transition period. Radio manufacturers and consultants are available to analyze existing links and to determine what changes are necessary to provide reliable performance. If the broadcast engineer is so inclined, there are many software analysis programs available to aid in link design.

Modulation Techniques

The FM system is one in which the frequency of the carrier is caused to vary in accordance with some

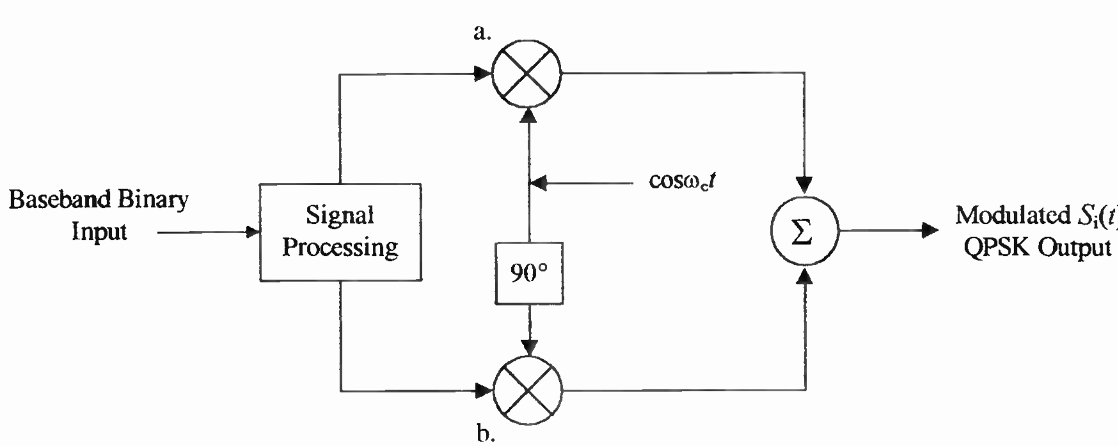


Figure 6.1-15. Simplified block diagram of QPSK modulator.

specified information carrying signal. For the simple case of a sinusoidal modulating signal at frequency f_m, the corresponding frequency modulated signed is given by $f(t)$ where:

$$f(t) = a \cos \omega_m t$$

where, "a" is the fixed amplitude of the modulated signal, and $\omega_m(\omega_m = 2\pi f_m)$ is the rate of modulation of the modulating signal.

Simply stated, there is a fixed amplitude sine wave of varying angular frequency ω_m. In the FM system all of the information is encoded as a function of the continuously varying angular frequency ω_m.

In digital systems, the amplitude, frequency and phase properties of the RF are quantized by the modulating signal. Digital implies a fixed set of discrete values. A digital radio waveform, then, can assume one of a discrete set of amplitude levels, frequencies or phases as a result of the modulating signal.

Frequency shift keying (FSK) consists of shifting the frequency of a sinusoidal carrier from a mark frequency (corresponding to a binary 1) to a space frequency (corresponding to a binary 0). This can be expressed as:

$$f_{c1}(t) = a \cos \omega_1 t$$

$$f_{c0}(t) = a \cos \omega_2 t$$

where "a" is the fixed amplitude of the modulated signal, a binary one corresponds to frequency f_1 and a binary zero to frequency f_2. An alternative representation of an FSK waveform consists of letting $f_1 = f_c - \Delta f$ and $f_2 = f_c + \Delta f$. The two frequencies then differ by $2\Delta f$ hertz and thus we can write:

$$f_c(t) = a \cos(\omega_c \pm \Delta\omega)t$$

(where $\Delta w = 2\Pi\Delta f$)

The frequency then deviates $\pm\Delta f$ about f_c. Δf is commonly called the *frequency deviation.*

Phase shift keying (PSK) consists of shifting the phase of a sinusoidal carrier 180° with a unipolar binary signal. This can be expressed as:

$$f_c(t) = \pm\cos \omega_c t$$

Here a binary one in the baseband binary stream corresponds to positive polarity and a binary zero to negative polarity. This modulation is commonly called *binary phase shift keying* (BPSK).

For BPSK, the values of 0° and 180° are typically used. These values are used to maximize (or to separate) the decision distance. There is no reason why 45° cannot be assigned as the binary 1 and 225° as the binary zero, as long as the distance between the two states is maximized.

Continuing along this line, instead of using two phase states, four can be used, each separated by 90°. The binary values to each of the four phase states can now be assigned as: 0° = 0, 1; 90° = 0, 0; 180° = 1, 1 and 270° = 1, 0. This type of multisymbol signaling is commonly called *quaternary phase shift keying* (QPSK) modulation or 4-PSK modulation. It can be expressed as follows:

$$S_i(t) = a_i \cos \omega_c t + b_i \sin \omega_c t$$

A simplified block diagram of a QPSK modulator is shown in Figure 6.1-16.

It is useful to represent the signal of $S_i(t)$ in a two-dimensional diagram locating the various points (a_i, b_i). The horizontal axis corresponding to the location of a_i is called the *inphase axis.* The vertical axis, along which b_i is located, is called the *quadrature axis.* The four signals of $S_i(t)$ assigned above then appear as shown in Figure 6.1-16. The points are said to represent a signal constellation, or as it is sometimes called, a *signal state diagram.*

More general types of multisymbol signaling schemes may be generated by letting a_i and b_i take on multiple values themselves. The resultant signals are called *quadrature amplitude modulation* (QAM) signals. These signals may be interpreted as having multilevel amplitude modulation applied independently on each of the quadrature carriers. A simplified block diagram of a QAM modulator is shown in Figure 6.1-17. The output of such a modulator can be expressed as:

$$S_i(t) = r_i \cos (\omega_c t + \theta_i)$$

where the amplitude r_i and the phase angle θ_i are given by the appropriate combinations of (a_i, b_i).

The signal constellations for a 4-, 16- and 64-QAM modulation is shown in Figure 6.1-18.

After careful examination of the PSK and QAM signal constellations one begins to appreciate the complexity required in a modem that can resolve the amplitude and phase information in multisymbol signaling schemes. One has to be more impressed with the modem that can resolve this information when the digital radio system is subject to external noise factors. To get a feel of how different digital modulations compare in the presence of thermal noise we can look at the bit energy to noise density ratio:

$$E_b/N_o$$

where, E_b is the energy per bit and N_o is the single sided noise spectral density. E_b/N_o can be expressed

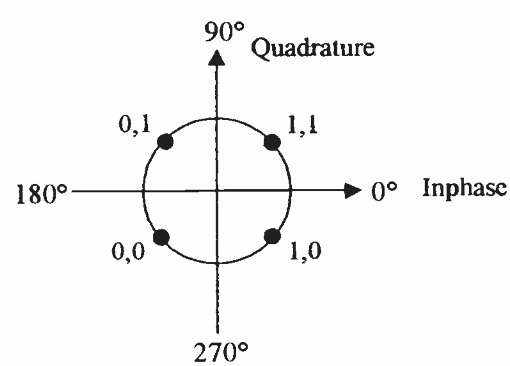


Figure 6.1-16. QPSK signal constellation.

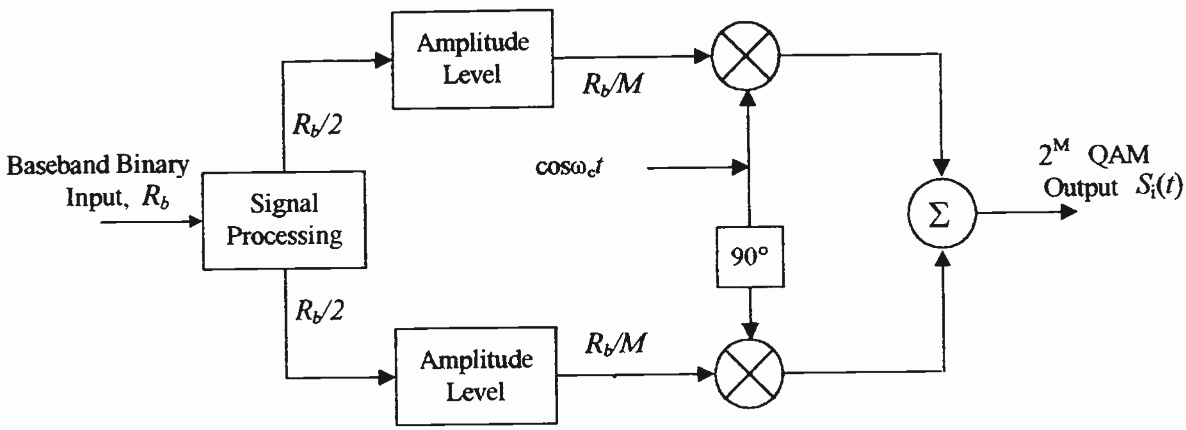


Figure 6.1-17. Simplified block diagram of a QAM modulator.

as the ratio of receive signal level (RSL) to the bit rate; in terms of log arithmetic, this can be written as:

$$E_b = \text{RSL}_{\text{dBm}} - 10\log(\text{bit rate})$$

The single sided noise spectral density of a perfect receiver in a 1 Hz bandwidth with a noise figure (NF) can be expressed as:

$$N_o = -174\text{ dBm} + \text{NF}_{\text{dB}}$$

The expression for the bit energy to noise density ratio is:

$$E_b/N_o = (\text{RSL}_{\text{dBm}} - 10\log(\text{bit rate})) - (-174\text{ dBm} + \text{NF}_{\text{dB}})$$

$$E_b/N_o = \text{RSL}_{\text{dBm}} - 10\log(\text{bit rate}) + 174\text{ dBm} - \text{NF}_{\text{dB}}$$

The above expression shows that as RSL increases so does the E_b/N_o. In other words, the greater that E_b/N_o needs to be, the less sensitive the receiver will be. This will be an important factor when determining the type of modulation to employ when converting existing analog links to digital.

Table 6.1-2 shows the E_b/N_o ratios based on a bit error rate (BER) of 10^{-6} for some common digital modulation formats and their bandwidth efficiencies based on the Nyquist bandwidth[1] (bandwidth numerically equal to the bit rate).

The parameters listed in Table 6.1-2 are idealistic because the only source of errors is due to thermal noise in the receiver. No modulation implementation loss is considered.

With the information given in Table 6.1-2 and from the E_b/N_o equation above, the receiver threshold for a T3 (45 Mbps) 16-QAM radio system with a radio noise figure of 4 dB can be calculated.

If we solve the E_b/N_o equation for $\text{RSL}_{(\text{min dBm})}$ we have the expression:

$$\text{RSL}_{(\text{min dBm})} = E_b/N_o + 10\log(\text{bit rate}) - 174\text{ dBm} + \text{NF}_{\text{dB}}$$

$$\text{RSL}_{(\text{min dBm})} = 17.0\text{ dB} + 10\log(44.736 \times 10^6) - 174\text{ dBm} + 4\text{ dB}$$

$$\text{RSL}_{(\text{min dBm})} = -76.5\text{ dB}$$

For a 64 QAM system the $\text{RSL}_{(\text{min dBm})} = -71.0$ dBm
For a 256 QAM system the $\text{RSL}_{(\text{min dBm})} = -65.7$ dBm

The example clearly shows that for higher order modulation schemes a greater E_b/N_o is required to enable the demodulator to extract the phase and amplitude information from the modulated signal. This inherent digital radio feature also causes the system gain to decrease as the modulation complexity increases. The previous example limits the distance between the transmitter and receiver as a function of the modulation scheme.

As a means to increase system gain, modem designers employ error correcting codes (or forward error correction (FEC)) to help improve performance. Error correcting coding usually requires redundancy (increasing the bit rate) and therefore, poses a contradictory requirement to obtaining maximum spectral efficiency. There is no set standard for the amount of FEC a manufacturer will employ in a given modem. During the modem design phase, trade-offs are made concerning BER performance, dispersive fade margin, spectral efficiency and cost.

For the applications being discussed in this chapter, satellite modems require more FEC than terrestrial modems to improve the system performance at threshold. Satellite links cover great distances and thermal noise becomes the dominant impairment. The composite bit rate at the output of a satellite modulator can be as much as 50% higher than the data rate into the modulator and as much as 10 dB of coding gain can be realized.

[1] The Nyquist bandwidth, in this context, represents the minimum theoretical bandwidth required to transmit the modulated signal if a brick wall filter were used in the modulation process. As a practical matter, the actual bandwidth required is always greater than this.

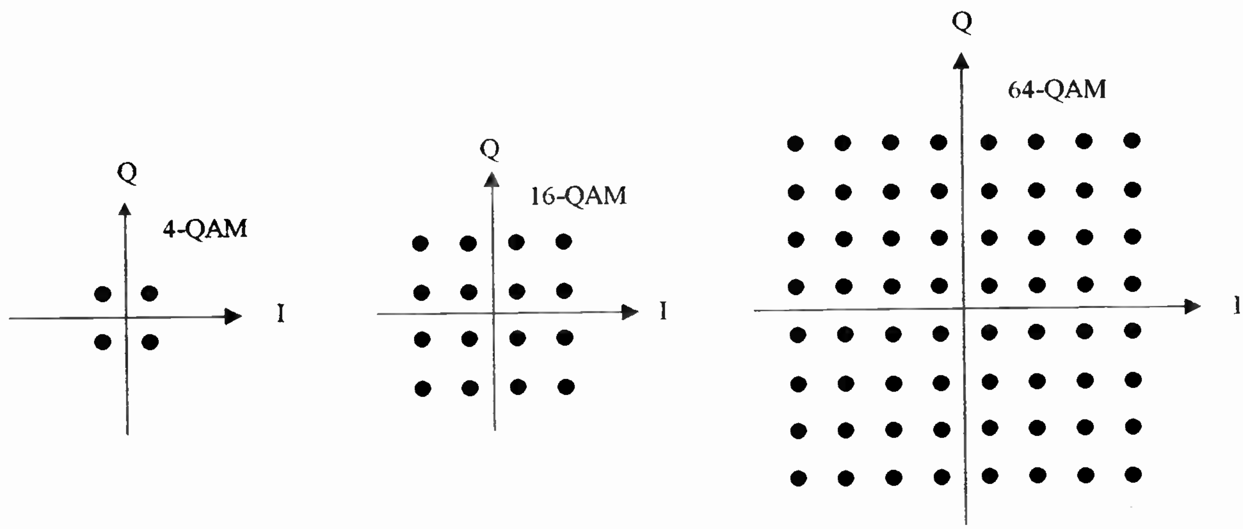


Figure 6.1-18. Signal constellations for a 4-, 16-, and 64-QAM.

A well designed terrestrial digital link is not fighting thermal noise as its major impairment. A modem designed for terrestrial use has to operate in a more hostile dispersive environment than the vacuum of space. The terrestrial modem designer has to balance coding gain, dispersive fade margin, spectral efficiency and cost. Terrestrial modems typically employ 5 to 15% of error correction redundancy. Table 6.1-3 shows the effect that 6.7 % of error correction redundancy has on QAM modulation (compare these values with the corresponding values shown in Table 6.1-2).

Again, FEC is not a cure-all for digital radio performance. The amount of error correction redundancy is up to the discretion of the modem designer and is just one of many factors that needs to be optimized during the modem design.

Transmitter

The transmitter in the digital radio link is the device that upconverts the modulated signal, amplifies it and then delivers it to the antenna system for transmission into free space. A simplified block diagram of a digital radio transmitter is shown in Figure 6.1-19.

As shown in Figure 6.1-19, an IF signal is first generated by the modulation portion of the system. The IF frequency is typically 70 or 140 MHz. This

Table 6.1-2
Basic Modulation Schemes

System	Variants	E_b/N_o (dB)	S/N (dB)	Nyquist Bandwidth
FSK	2-state FSK with discriminator detection	13.4	13.4	B
	3-state FSK (duo-binary)	15.9	15.9	B
	4-state FSK	20.1	23.1	B/2
PSK	2-state PSK with coherent detection	10.5	10.5	B
	4-state PSK with coherent detection	10.5	13.5	B/2
	8-state PSK with coherent detection	14.0	18.8	B/3
	16-state PSK with coherent detection	18.4	24.4	B/4
QAM	16-QAM with coherent detection	17.0	20.5	B/4
	32-QAM with coherent detection	18.9	23.5	B/5
	64-QAM with coherent detection	22.5	26.5	B/6
	128-QAM with coherent detection	24.3	29.5	B/7
	256-QAM with coherent detection	27.8	32.6	B/8
	512-QAM with coherent detection	28.9	35.5	B/9

Table 6.1-3
QAM modulation schemes with FEC added.

System	Variants	E_b/N_o (dB)	S/N (dB)	Nyquist Bandwidth
QAM	16-QAM with coherent detection	13.9	17.6	B/4 x (1 + .067)
	32-QAM with coherent detection	15.6	20.6	B/5 x (1 + .067)
	64-QAM with coherent detection	19.4	23.8	B/6 x (1 + .067)
	128-QAM with coherent detection	21.1	26.7	B/7 x (1 + .067)
	256-QAM with coherent detection	24.7	29.8	B/8 x (1 + .067)
	512-QAM with coherent detection	25.8	32.4	B/9 x (1 + .067)

signal is then upconverted by an image reject mixer. This translation of the modulated IF signal up to the RF carrier frequency is made by a mixing operation or multiplication between two frequencies. The result of this mixing operation is a summation of frequency components that can be expressed as:

$$nF_{LO} \pm mF_{IF}$$

where $n = 1, 2, 3, \ldots$ and $m = 1, 2, 3, \ldots$.

The frequencies that are of most concern for the case when $n = m = 1$, or:

$$F_{LO} + F_{(IF)}$$

and

$$F_{LO} - F_{(IF)}$$

The desired frequency for transmission is:

$$F_c = F_{co} + F_{(IF)}$$

$F_{LO} - F_{(IF)}$ is the difference frequency, which is commonly called the *image frequency*. The image frequency is typically suppressed 20 to 25 dB by the image rejection properties of the mixer.

In addition to the sum and difference frequencies, there will be some LO leakage present at the output port of the mixer. The LO leakage is attenuated approximately 20 dB by the mixer's LO to RF isolation. Some digital transmitter designs may employ a bandpass filter to remove the image and LO leakage signals.

The local oscillator that drives the mixer in a digital radio is a high stability, low phase noise frequency source. The low phase noise LO is required when implementing multisymbol modulation. Phase noise (or short-term instability, or jitter) is seen near the carrier and causes degradation of BER. As can be seen in the signal constellation diagrams shown in Figure 6.1-18, LO phase noise causes the constellation points to jitter about. If the LO in the transmitter and receiver were an ideal frequency source the constellation points would not move at all. The more phase noise the LO induces onto the modulated IF signal, the more difficult signal recovery becomes. Figure 6.1-20 shows the empirically derived phase noise requirements for QPSK, 16-QAM and 64-QAM modulation. Notice that as the modulation complexity increases the LO phase noise needs to decrease.

Continuing the journey through the transmitter, the signal out of the frequency translating mixer is input into a linear operating amplifier (typically solid-state). A linear operating amplifier is one that is operating below its 1 dB gain compression point. The 1 dB gain compression (called G_{1dB}) is defined as the power gain where the amplifier's non-linearity reduces the power gain by 1 dB over the small signal linear power gain. That is,

$$G_{1dB}(\text{dB}) = G_o(\text{dB}) - 1$$

where G_o(dB) is the small signal linear power gain in decibels. Since the power gain is defined as

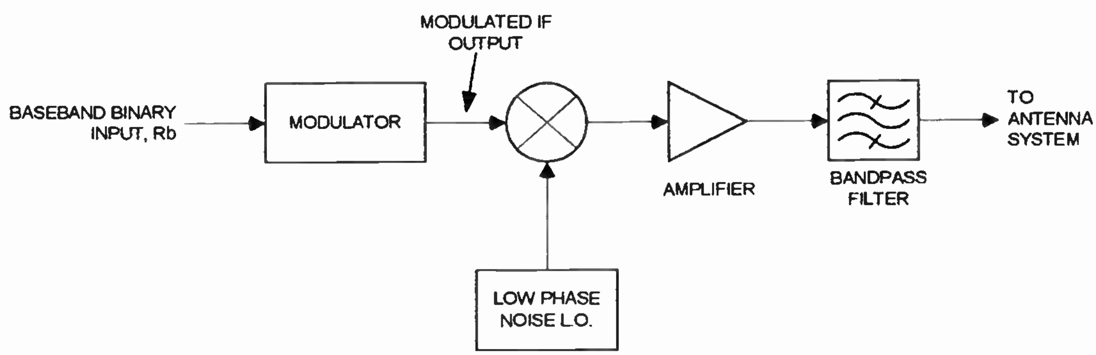


Figure 6.1-19. Simplified block diagram of a digital transmitter.

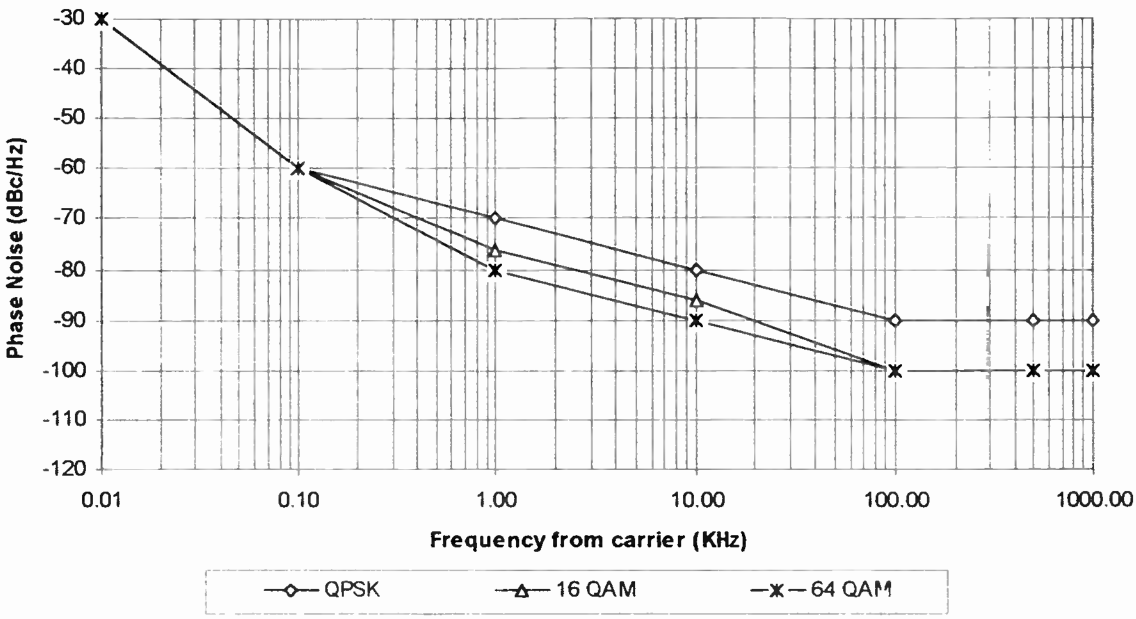


Figure 6.1-20. LO phase noise required for QPSK, 16-AQM and 64-QAM modulation.

$$G_p = P_{out}/P_{in}$$

or

$$P_{out}(\text{dBm}) = G_p(\text{dB}) + P_{in}(\text{dBm}).$$

The output power at the 1 dB gain compression point, called P_{1dB}, is

$$P_{1dB}(\text{dBm}) = G_{1dB}(\text{dB}) + P_{in}(\text{dBm}).$$

Substituting $G_{1dB}(\text{dB}) = G_o(\text{dB}) - 1$, into the above equation gives

$$P_{1dB}(\text{dBm}) - P_{in}(\text{dBm}) = G_o(\text{dB}) - 1.$$

The equation shows that the 1 dB gain compression point is that point at which the output power minus the input power in dBm is equal to the small signal power gain minus 1 dB. A typical plot of P_{out} versus P_{in}, illustrating the 1 dB gain compression point is shown in Figure 6.1-21.

In the digital radio transmitter that employs linear modulation, distortion can be caused by the power amplifier operating near or beyond $G_{1dB}(\text{dB})$. This distortion is caused by intermodulation (IM) products that arise when two or more sinusoidal frequencies are applied to a non-linear amplifier. The output of the amplifier will contain additional frequencies called *intermodulation products.* For example, if two sinusoidal signals

$$v(t) = a \cos 2\pi f_1 t + b \cos 2\pi f_2 t$$

are applied to a nonlinear amplifier whose output voltage can be represented by the power series

$$v_o(t) = \alpha_1 v(t) + \alpha_2 v^2(t) + \alpha_3 v^3(t)$$

the output signal will contain frequency components at dc, f_1, f_2, $2f_1$, $2f_2$, $3f_1$, $3f_2$, $f_1 \pm f_2$, $2f_1 \pm f_2$, and $2f_2 \pm f_1$. The frequencies $2f_1$ and $2f_2$ are the second harmonics, $3f_1$ and $3f_2$ are the third harmonics, $f_1 \pm f_2$ are the second order intermodulation products (since the sum of the f_1, and f_2 coefficients is 2), $2f_1 \pm f_2$, and $2f_2 \pm f_1$ are the third order intermodulation products (since the sum the f_1, and f_2 coefficients is 3, and so on). The input and output power spectra for a typical solid-state amplifier is shown in Figure 6.1-22.

Figure 6.1-22 shows that the third-order intermodulation products at $2f_1 - f_2$ and $2f_2 - f_1$ are very close to the fundamental frequencies f_1, and f_2 and (typically) fall (typically) within the amplifier bandwidth, producing distortion in the output.

The digital radio system that employs linear modulation and operates the power amplifier near or into

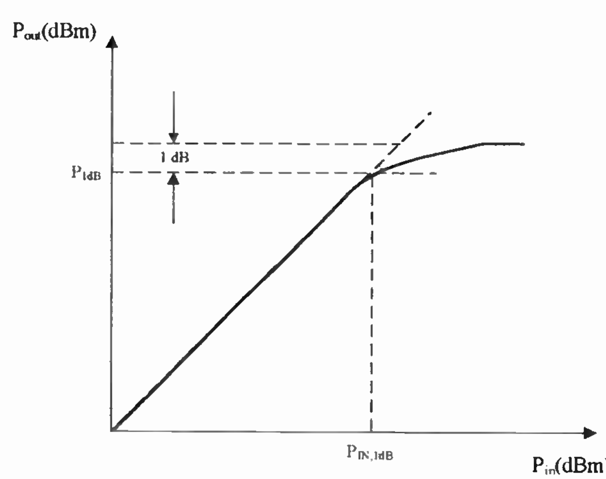


Figure 6.1-21. The 1 dB gain compression point.

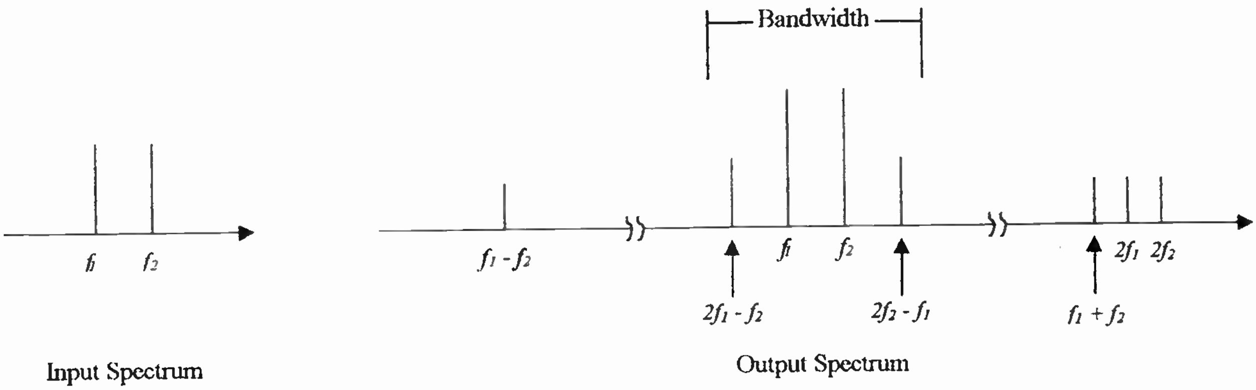


Figure 6.1-22. Input and output power spectrum.

saturation (beyond G_{1dB}(dB)) will see distortions that are caused by the third order IM products. The distortions can be seen as a spreading of the spectrum that can occupy up to three times the bandwidth of the original spectrum. Third order IM distortion will degrade BER performance and can cause interference to adjacent channels.

The severity of third order IM distortion is a function of where the amplifier output level is operating in relation to its P_{1dB} point. In addition, the more complex the modulation scheme, the larger the required P_{1dB} back off to minimize IM distortion. Table 6.1-4 shows the typical power amplifier back offs for FSK, QPSK and QAM modulation.

Figure 6.1-23 shows the spectrum plots of a 16-QAM 70 MHz modulated signal at the modem output and the same signal upconverted to 7,037.8 MHz at the transmitter output. In this case the transmitter power amplifier was backed off by 6 dB from P_{1dB}. It is important to note that even at this operating point, some spreading still occurs.

Finally, the amplified signal is bandpass filtered and directed to the antenna system for free space transmission. At the far end of the free space path the modulated microwave signal is captured by the receive antenna system.

Receiver

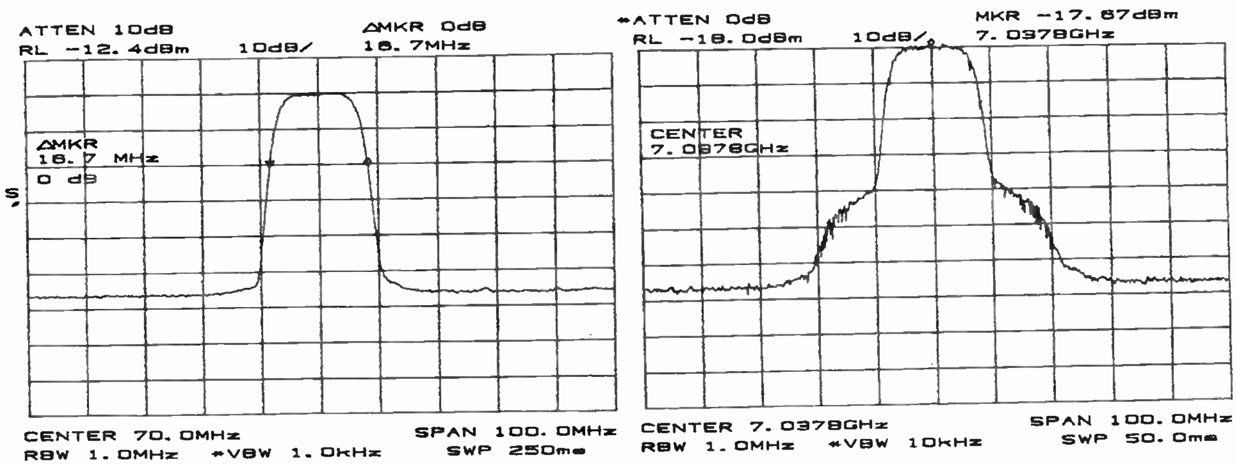
The receiver in the digital radio link is the device that receives the modulated signal from the antenna system, filters it, amplifies it, downconverts it to IF, filters and amplifies it again and then delivers it to the demodulator for signal processing back to baseband. A simplified block diagram of a digital radio receiver is shown in Figure 6.1-24.

In the digital link, the receiver performs the lion's share of the work. In Figure 6.1-18, the 4-, 16- and 64-QAM signal constellations were shown. The figure shows that as the value of M (M = 4, 16, 64) increases, the space between the constellation points decreases. For the digital link to operate, the digital receiver must be capable of resolving which signal point was transmitted. For 64-QAM modulation, it has to resolve 64 points. For 512-QAM modulation, it has to resolve 512 points!

All of the transmitter issues, previously discussed concerning LO phase noise and amplifier linearity, hold true for the digital receiver. In addition to these internal concerns, the receiver is also susceptible to

Table 6.1-4
Typical power amplifier P_{1dB} bak off.

System	Variants	Typical Back-off from P_{1dB} (dB)
FSK	2-state FSK with discriminator detection	0
	3-state FSK (duo-binary)	0
	4-state FSK	0
PSK	2-state PSK with coherent detection	1
	4-state PSK with coherent detection	2
	8-state PSK with coherent detection	4
QAM	16-QAM with coherent detection	6
	64-QAM with coherent detection	10
	256-QAM with coherent detection	12



16-QAM signal at the modulator output

16-QAM signal at the transmitter output

Figure 6.1-23. The RF spectrum of 16-QAM signal at the modulator output and transmitter output.

outside disturbances that tend to deteriorate system performance. It is an understatement to say that the receiver must be capable of resolving which signal point was transmitted. The receiver must be capable of resolving which signal point was transmitted in the presence of transmitter distortions, receiver distortions, white Gaussian noise, co-channel and/or adjacent channel interference and multipath distortions.

The transmitter and receiver internal distortions can be reduced by employing low phase noise local oscillators and operating internal amplifiers sufficiently below their P_{1dB} point.

Additive white Gaussian noise is unavoidable. The bandwidth efficient modem requires a higher E_b/N_o for a given symbol rate as the number of bits per second per Hertz is increased.

Co-channel and/or adjacent channel interference in most cases can be avoided by frequency allocation management, channel filtering and meeting spectral masks.

Multipath distortions are caused by the transmission channel itself. To minimize dispersive distortions the broadcast engineer must design the link to avoid ground reflections, implement space and/or frequency diversity and employ adaptive equalization.

System Considerations

The implementation of digital video line-of-sight (LOS) radio links is accelerating in the broadcast industry primarily due to the emergence of MPEG-2 technology. Once the broadcast engineer overcomes the fear of change, they will realize the advantages that a well designed digital link offers. In digital radio systems, thermal and intermodulation noise accumulation can to a large extent be disregarded over the network. Noise becomes a more isolated problem between points of regeneration. This is an overwhelming advantage over analog transmission, where the primary concern of the broadcast engineer is noise accumulation.

The unit of digital radio performance is (BER rather than S/N noise accumulation) which was the measures for analog radio link design. The BER is the ratio of the number of *errored* bits received to the *total* number of bits received over a given time interval. In digital telephony systems, the BER threshold is approximately $10e^{-3}$; this was established around the distribution of voice data. In digital video radio systems, the BER threshold is approximately $10e^{-6}$; this criterion is established around the video encoding and decoding equipment.

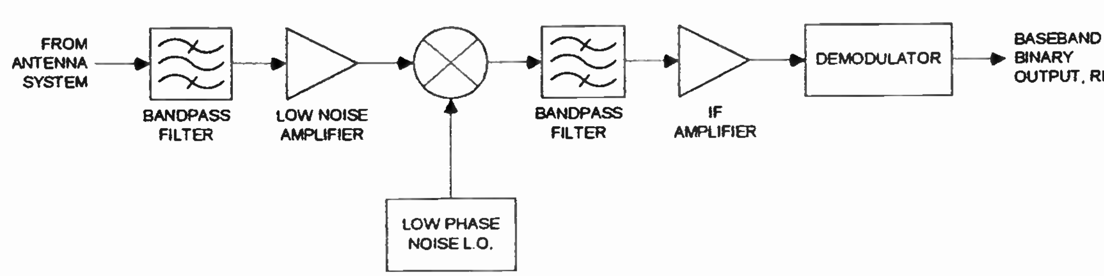


Figure 6.1-24. Simplified block diagram of a digital receiver.

Fade Margin

For a well designed LOS microwave path that is not subjected to diffraction fading or surface reflections, multipath conditions provide the dominant factor for fading below 10 GHz. Above 10 GHz, the effects of rainfall determine maximum path lengths through system availability criteria.

Digital radios react to fading differently than their analog counterparts. Analog systems are more concerned with a thermal fade margin. Digital systems are more concerned with the dispersiveness of a path. How well the radio operates in a dispersive environment is call the *dispersive fade margin*. In a digital video radio system the fade margin consists of four factors that are power added to give the *composite fade margin* (CFM) (sometime referred to as *effective fade margin*).

1. *Thermal fade margin (TFM)* (or flat fade margin) is the algebraic difference between the normal RSL and the $10e^{-6}$ BER outage threshold for flat non-dispersive fades.
2. *Dispersive fade margin (DFM)* also referenced to the $10e^{-6}$ BER, is defined by the radio equipment manufacturer. It is determined by the type of modulation, the effectiveness of the adaptive equalization employed in the receive path (if used) and the multipath signal delay time. For purposes of characterization, the multipath delay time has been standardized to 6.3 ns. The manufacturer characterizes a modem by introducing a 6.3 ns frequency response notch into the IF passband. The center of the notch is stepped across the passband. At each frequency step, the depth of the notch is adjusted to produce a $10e^{-6}$ BER. The result of a typical radio response is shown in Figure 6.1-25. The response shown is referred to as the *equipment signature, M* or *W curves*. The DFM characterizes the digital radio's robustness to dispersive fades; a DFM greater than 50 dB is a good baseline criterion.
3. *External interference fade margin (EIFM)* is the receiver threshold degradation due to external system, independent of thermal noise.

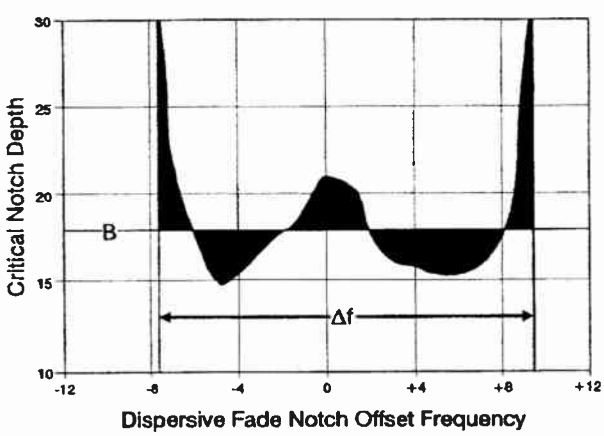


Figure 6.1-25. Dispersive fade margin measurement.

4. *Adjacent channel interference fade margin (AIFM)* is the receiver threshold degradation due to interference from adjacent channel transmitters in one's own system. This is normally a negligible parameter except in the cases of frequency diversity and multiline hot standby systems.

The four fade margins are power added to derive the CFM as follows:

$$\text{CFM} = 10\log(10^{-\text{TFM}/10} + 10^{-\text{DFM}/10}\,10^{-\text{EIFM}/10} + 10^{-\text{AIFM}/10})$$

The outage time due to multipath fading is calculated by:

$$T = (rT_o \times 10^{-(\text{CFM}/10)})/I_o$$

where

T = outage time in seconds
r = fade occurrence
T_o = $(t/50)(8 \times 10^6)$ = length of fade season in seconds
t = average annual temperature in degrees Fahrenheit
CFM = composite fade margin
I_o = space diversity improvement factor; factor = 1 for no diversity; ≥ 1 for space diversity.

The fade occurrence factor, r, is calculated from the basic outage equation for atmospheric multipath fading:

$$r = c(f/4)(D/1.6)^3 \times 10^{-5} \quad \text{(English)}$$

$$r = c(f/4)D^3 \times 10^{-5} \quad \text{(metric)}$$

where

c = climate or C factor (see Figure 27)
f = frequency (GHz)
D = path length (miles or km)

The space diversity improvement factor I_o may be calculated by:

$$I_o = 7 \times 10^{-5} s^2(f/D) \times 10^{\text{CFM}/10}, s \leq 50\text{ ft (English)}$$

$$I_o = 1.2 \times 10^{-3} s^2(f/D) \times 10^{\text{CFM}/10}, s \leq 15\text{ m (metric)}$$

where, s, is the vertical antenna separation in meter (feet), center to center.

The antenna spacing rule of thumb in a diversity system is a minimum of 200 wavelengths. For example, one wavelength at 6 GHz is 5 cm. The required separation, then would be 5 × 200 or 10 m (approximately 33 ft).

The climate or "C" factor shown in Figure 6.1-26, is a constant that is used in the fade occurrence equation that emulates climate and terrain effects on propagation. Mountain and dry climates are areas of good propagation conditions and have a C factor of 0.25. Paths over or near water and gulf coastal areas are of difficult propagation conditions and have a C factor of 6.

The following system parameters of an actual digital video microwave link show the different climate and terrain regions of the United States affect digital radio performance:

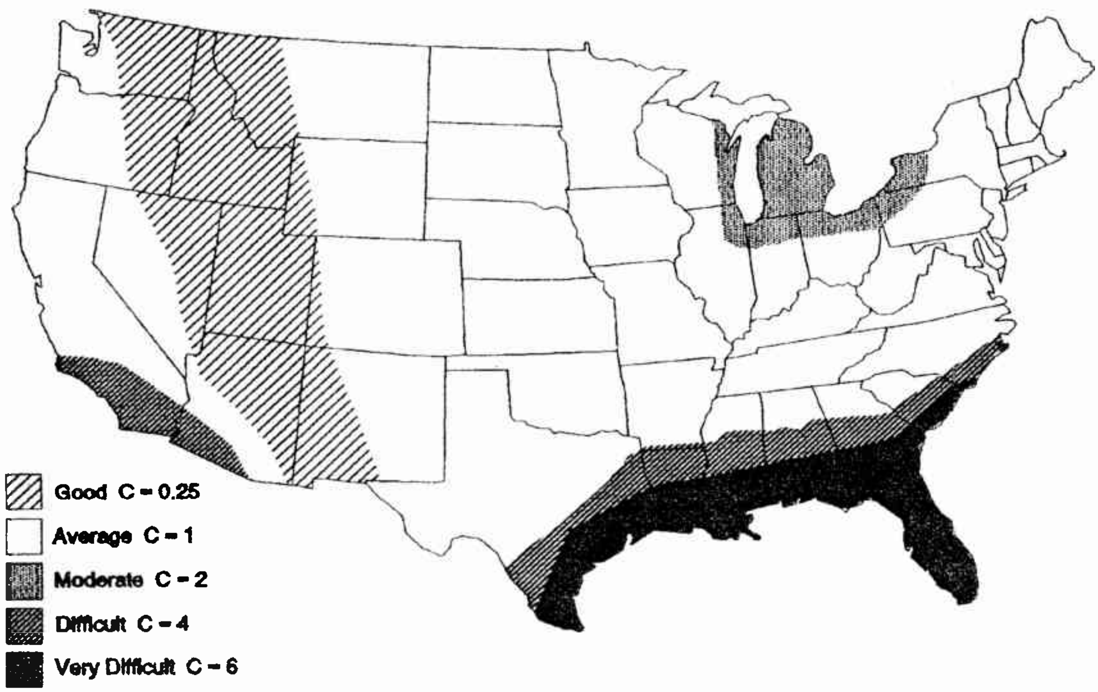


Figure 6.1-26. Values of climate factor for the continental United States.

Transmit frequency: 7,012.5 MHz
Modulation type: 16-QAM
Data rate: 44.736 Mbps
Transmit power: +27 dBm
Receiver threshold: −80 dBm at a BER of $10e^{-6}$

Dispersive fade margin: 50 dB
External interference fade margin: negligible
AIFM: negligible
Annual average temperature: varies according to Figure 6.1-27

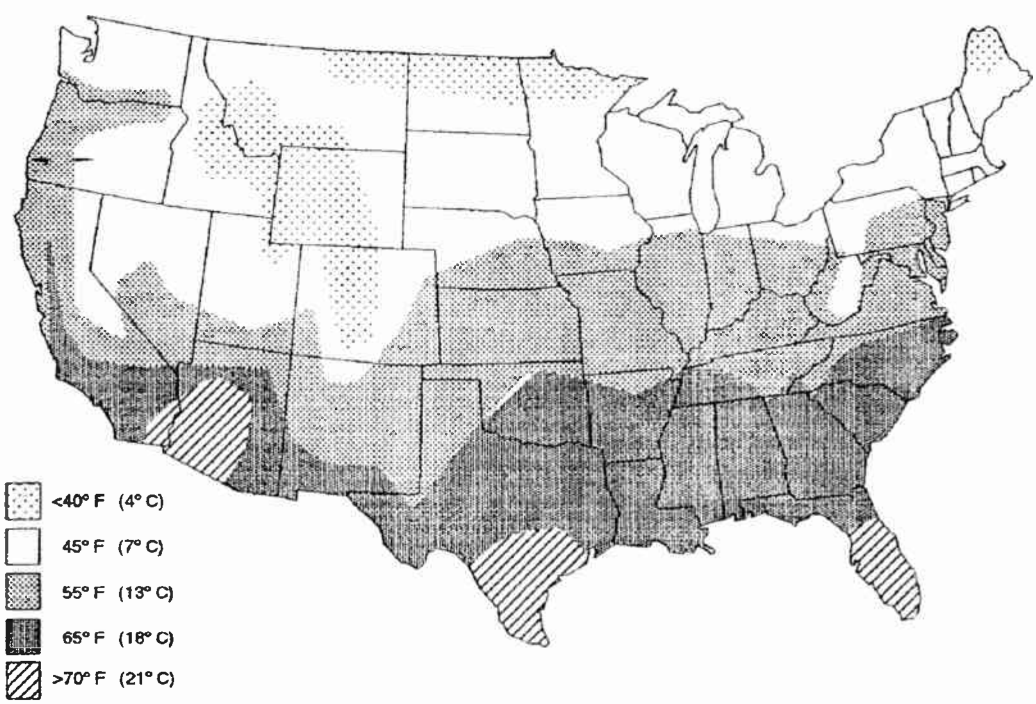


Figure 6.1-27. Average annual temperature for the United States.

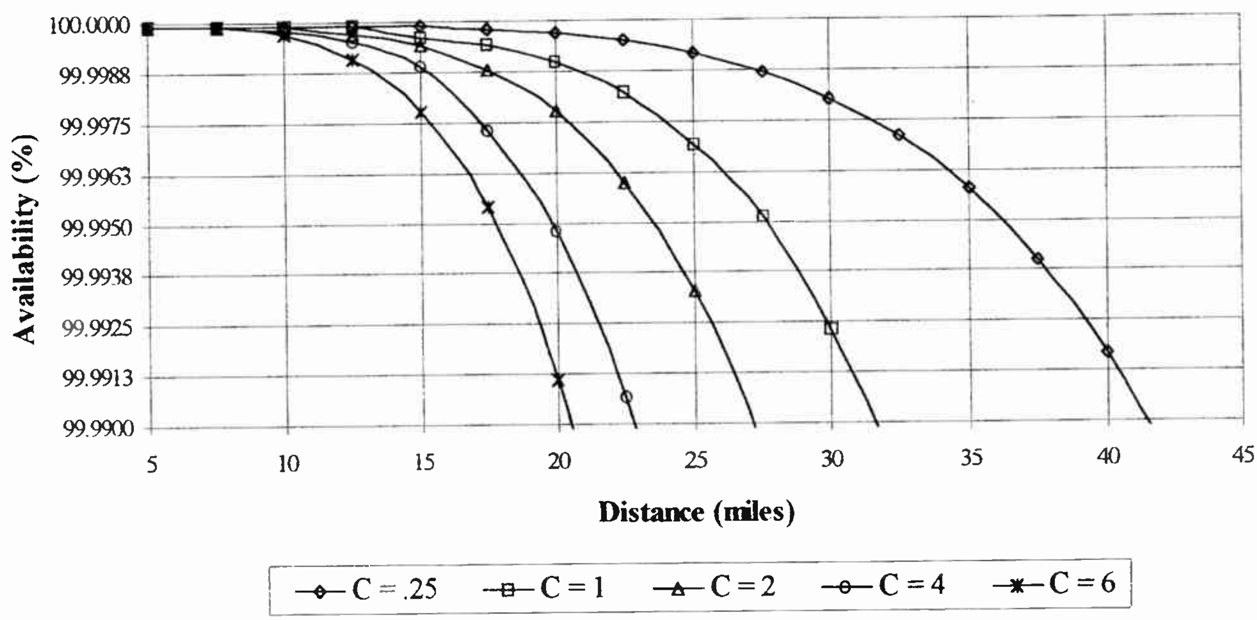


Figure 6.1-28. Link availability as a function of C factor and distance (no diversity).

Path length: varies from 5 to 50 miles
Antenna Diameter: 8 feet with a gain of 41.5 dB
Waveguide lengths: two 300-foot lengths at a loss of 1.4 dB per 100 ft.

With this system information the availability for this link as a function of path length and C factor can be calculated. For this example, assume no effects due to earth bulge or surface reflections. The result of this example is shown in Figure 6.1-28.

To improve the link availability over long paths and difficult propagation areas space diversity can be employed. Space diversity improvement plays such a significant role in increasing path reliability, that it often allows higher interference levels that degrade (reduce) the composite fade margin of many digital links.

Figure 6.1-29 shows the previous example with space diversity. The antenna spacing for this case was approximately 200 wavelengths (approximately 30 ft).

The space diversity improvement factor (I_o) may underestimate diversity improvements for small an-

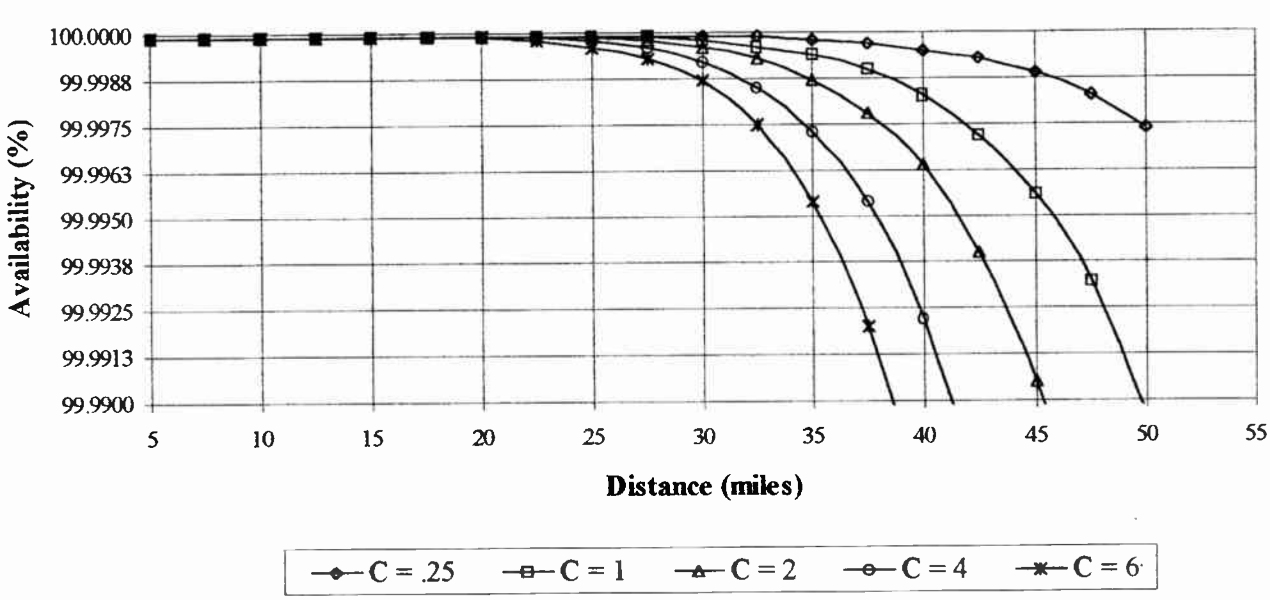


Figure 6.1-29. Link availability as a function of C factor and distance (with space diversity).

tenna spacings and overestimate diversity improvement for large antenna spacings on flat land microwave links.

Angle diversity is an alternative approach to space diversity when atmospheric dispersive fading is the issue. Angle diversity employs a monopulse sum and difference beam to achieve angular displacement. The monopulse beams are created by a hybrid antenna feed. The hybrid outputs are connected in a manner similar to a space diversity configuration. The sum beam is fed into a receiver and the difference beam is fed into a receiver. A typical angle diversity antenna system is shown in Figure 6.1-30.

If one beam exhibits multipath fading, the second beam, due to a vertical angular difference in pattern, will have a different signal relationship from the multipath ray and will have a stronger signal and less fading than the first. By switching between beams, the maximum fade on the path may be significantly reduced. This is true for dispersive multipath fading in digital radio systems that have sufficient thermal fade margin of at least 40 dB.

Digital Video LOS Microwave Repeaters

Digital LOS microwave repeaters completely demodulate the incoming signal to baseband (to the raw electrical signal of "1s" and "0"). This full demodulation permits regeneration of the "1s" and "0s." The outgoing signal is squared up and retimed. Regenerative repeaters are an essential key to digital transmission in that the noise stops at the repeater.

The regenerated baseband signal is remodulated, upconverted and transmitted on a different frequency. RF and IF repeaters are not recommended for digital microwave systems because such repeaters do not have the vital regeneration stage.

PRIMARY POWER & BACKUP

With improvements in equipment and systems, engineering the reliability of the primary power source becomes a major factor in overall system availability. Where the terminals are in studios or TV transmitter sites the ac supplies may well be protected with standby generators; in this case, ac sourcing is the obvious choice. At repeaters, the ac may be supplied to a remote site by overhead lines and these may be subject to interruption under adverse weather conditions. Since solid-state microwave equipment has essentially low power requirements, batteries are the most popular form of standby power. Where long outages are to be expected (12 hours or more) an ac generator will be needed to recharge the batteries: this can be fuelled by gasoline, diesel or liquified natural gas (propane). In addition, the use of solar power should be considered.

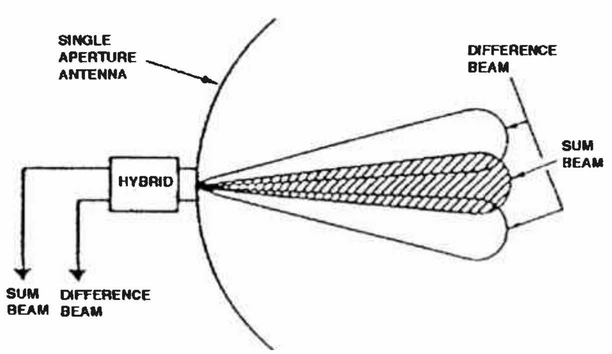


Figure 6.1-30. A typical angle diversity antenna system.

Industrial lead-acid batteries are the most common type. They can be sealed to prevent evaporation of liquid, reducing maintenance and the need to ventilate the explosive gas given off when a battery is charged at too high a rate. The cells are floated across the charger, which supplies the station load until the ac fails. Since different types of lead-acid cell have different float voltages (2.15, 2.2 and 2.23 VDC for lead-antimony, lead-calcium and lead-selenium respectively) it is important that the charger and battery are matched—for this reason it is recommended that both are ordered from the same supplier.

It must be remembered that a battery of 12 cells with a nominal voltage of 24 VDC will float at 26.8 VDC, dropping to 24 when the charging fails, and finally falling to 21 VDC at the end of the standby time. At a minimum, the electronic equipment must accept this range. To determine the capacity of each cell required to give a particular standby time (H hours) the steady current drain must be determined (A amperes). Depending on the type of regulation used in the electronic equipment this may be higher at 26.8 VDC or 21 VDC (usually the latter)—take the higher value.

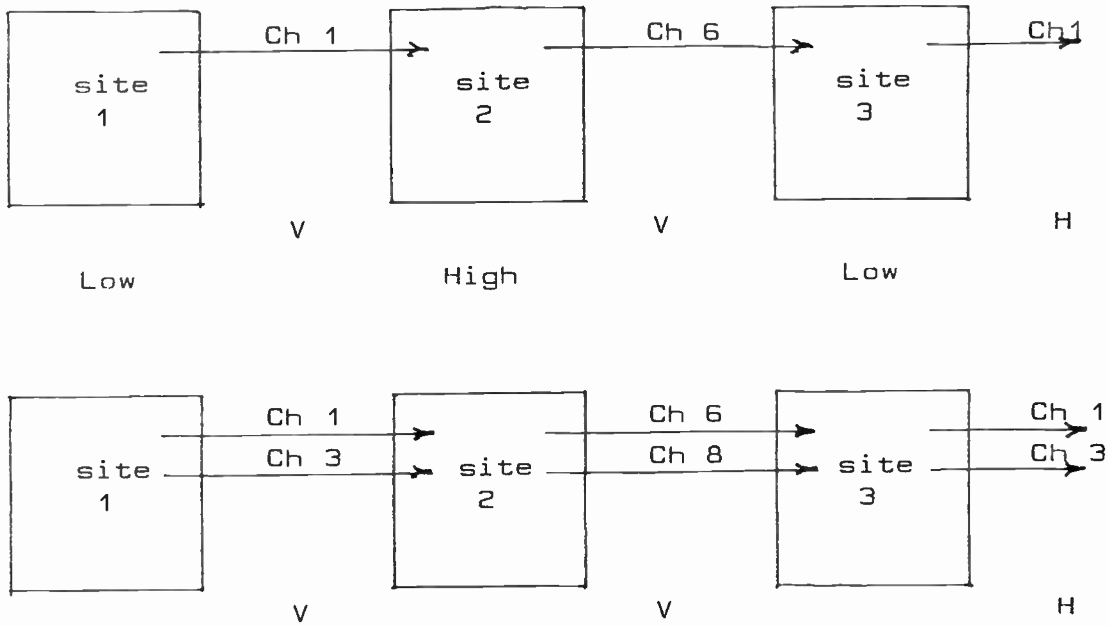
To arrive at the required cell capacity, multiply A by H to get the capacity in ampere-hours (AH). The minimum size of charger (C amperes) to meet the station load and at the same time recharge a discharged battery in R hours is given by:

$$C = (A + AH) \times 1.1/R$$

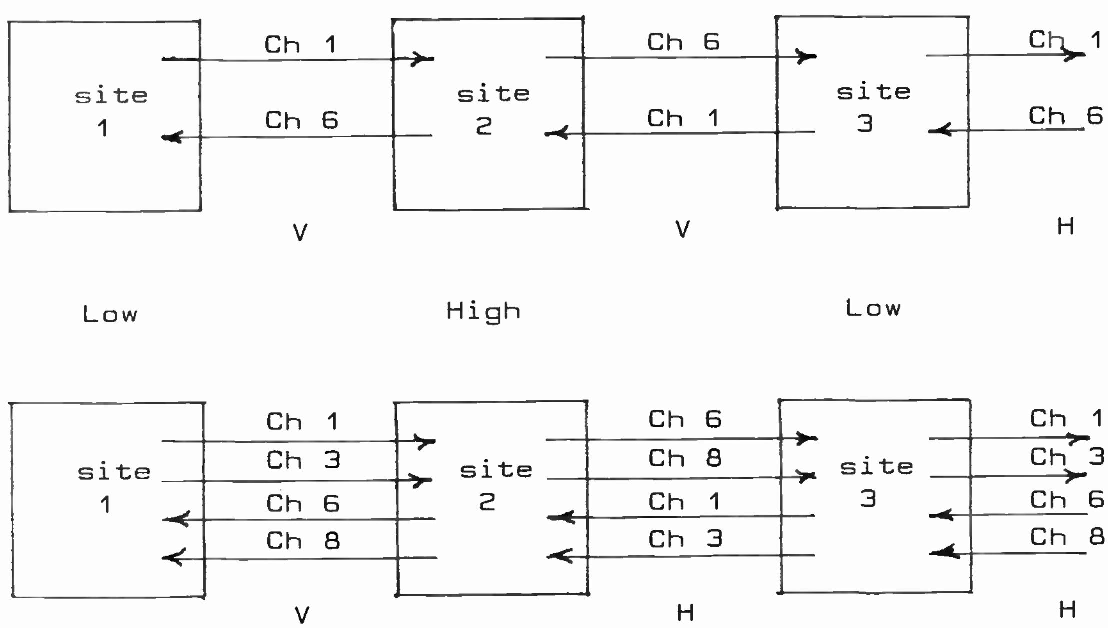
Since charger failure would lead to station failure after H hours it is common practice to use a duplicated charger for full protection at key sites; in that case each charger need only have the capacity A amperes rather than C as calculated above. The chargers must be designed to share the load.

The battery acts as a large capacitor, reducing the ripple voltage generated by the charger and protecting against powerline surges. A good charger will have low voltage protection to isolate the battery when discharged and over voltage protection and charge rate limiting to protect against surges and gassing by the cells. When diesel generators are used as the primary source of power, they should be run at 75% or more of their rating (after allowing for any derating for altitude).

Solar power using photovoltaic cells is attractive as there are no moving parts requiring maintenance. However, the batteries used to maintain the supply during hours of darkness or heavy overcast (which



A: One-channel and two-channel simplex systems



B: One-channel and two-channel duplex systems

Figure 6.1-31. Recommended initial and expanded frequency plans; channels 2, 4, 7, 9 can be added on the opposite polarisation, and channels 5, 10 can be used for spur routes.

could exist for several days in some parts of the country) can be the most expensive part of the installation. It may pay to have a small ac generator and charger to back up the batteries if the solar input fails for more than three days, rather than batteries to maintain the supply for six days, for example. Such a generator could also power lights and test equipment during routine site visits.

FREQUENCY PLANNING

Frequency coordination for the broadcast microwave industry has been primarily a local case-by-case process through cooperation by the respective chief engineers. Most areas, through the local chapter of the Society of Broadcast engineers (SBE), maintain an up-to-date log of existing licenses and proposed applications.

Ideally, frequency planning for future growth should follow certain simple ideas. Foremost among these is the concept of dividing each available band into *high* and *low* parts and trying to keep the transmitters at a given site in one or other part of the band. This ensures the maximum utilization of frequencies at every site. Adjacent channels should be cross polarized; second channels on a given route should be co-polarized on the next-but-one channel; if a *go* channel is on an odd channel then the corresponding *return* channel should be on an even channel to avoid intermodulation products (2A-B). The diagrams of Figure 6.1-31 illustrate this practice.

Frequency Re-use

Microwave antenna discrimination provides a degree of isolation that very often allows the reuse of the same channel at the same or a nearby site. On parallel routes where the terminals at each end are several hundred yards apart, it is often possible to use the same channel on both routes but in different directions. Every case has to be studied carefully to take account of the free space losses and the angular discrimination of the antennas.

TIA Telecommunications Systems Bulletin TSB 10-F *Interference Criteria for Microwave Systems* specifies a 60 dB co-channel carrier/interferer ratio for analog FM video systems. This is an ideal, but in many cases may not prove possible. Lowering the C/I ratio will raise the effective noise level and hence reduce the fade margin. Given a threshold of -85 dBm the noise level will be approximately -95 dBm and an interferer at -95 dBm will double the noise level and raise the threshold to -82 dBm. This, or an even higher level, may be acceptable in a difficult situation. The interferer must at all times be below the threshold of the receiver or capture will occur if the wanted signal fails.

FCC LICENSING

All broadcast microwave systems, whether for fixed or mobile applications, must be licensed by the FCC. The application is filed under Part 74 of the FCC Rules on Form 415 with pertinent characteristics and the Type Acceptance Identification Number provided by the manufacturer of the chosen equipment.

Frequency coordination for clearance of the requested frequencies is the responsibility of the local area engineers. In contrast, licensing by common carriers, industrial service and common antenna radio systems (CARS) band operators is supported by formal frequency coordination studies. Broadcast engineers should be aware and careful when applying for channels in the 12.7–13.25 GHz band. This band is shared with CARS and cable TV operators and is in wide use in most areas.

The details in Form 415 are straightforward and the instructions that are supplied by the FCC are very thorough. There is a main form and four schedules. The main form is required with all applications, as is Schedule A (Transmit Site Data). Note that the output power given in the Type Acceptance is a maximum; when quoting antenna input power and effective radiated power (ERP) the true transmitter power less any feeder losses (for input power) and plus the antenna gain (dBi) for ERP, should be quoted. Copies of Schedule B (Frequency Data) and Schedule C (Path Data) are required for each path in a multihop system. Schedule D (Passive Repeater Stations) is only needed when these are involved. The emission designator is normally provided by the equipment supplier.

Frequency agile portable transmitters are typically licensed for all the channels that they can provide. For example, at 2 GHz all seven channels are submitted on a single license application.

Broadcasters and others using this *Handbook* should refer to the FCC *Code of Federal Regulations* and to their attorneys for complete up-to-date information.

REFERENCES

Andrew Catalog 37, Andrew Corporation, Orland Park, IL., 1997.

Oster, J, and E. Bachner, *Angle Diversity: A Practical Technique for Reducing Fades,* SP20-45, Andrew Corporation, Orland Park, IL, 1989.

Cablewave Systems Catalogs 720C & 800, Radio Frequency Systems, Inc., Cablewave Systems Division, North Haven, CT, 1992.

C.C.I.R. Documents of the Seventeenth Plenary Assembly, Düsseldorf, Germany, ITU, Geneva, Switzerland, 1990.

K. Feher, *Digital Communications Microwave Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1981.

Effects of Multipath Propagation on the Design and Operation of Line-of-Sight Digital-Relay Systems, ITU-R Rec., F.1093, 1994 F Series Vol., Part 1, ITU, Geneva, 1994.

Electrical Performance Standards for Television Relay Facilities, Standard EIA/TIA 250-C, Telecommunications Industries Association, Washington, DC, 1990.

Interference Criteria for Microwave Systems, TIA Telecommunications Systems Bulletin TSB 10-F, Telecommunications Industries Association, Washington, DC, 1994.

Information Transmission, Modulation, and Noise, 3rd ed., Schwartz, McGraw-Hill, New York, NY, 1984.

Serafin, R., *LO Phase Noise Requirements for QSPK, 16-QAM and 64-QAM Modulation,* Microwave Radio Communications Memo, Microwave Communications, Chelmsford, MA, September 3, 1996.

Kolberg, Erik L., *Microwave and Millimeter-Wave Mixers, IEEE Press,* New York, NY, 1984.

Hickin, E. M., "Microwave Engineering for the Broadcaster," *NAB Engineering Handbook,* 8th ed., National Association of Broadcasters, Washington, DC, 1992.

Passive Repeater Engineering, Microflect Company, Inc., Salem, OR, 1989.

Pathloss, Contract Telecommunication Engineering Ltd., Coquitlam, BC, Canada.

Freeman, Roger L., *Radio System Design for Telecommunications,* 2nd ed., John Wiley & Sons, Inc., New York, NY, 1997.

Hogg, D. C., *Statistics on Attenuation of Microwaves by Intense Rain,* Bell System Technical Journal, November, 1969.

Freeman, Roger L., *Telecommunications System Engineering,* 3rd ed., John Wiley & Sons, Inc., New York, 1996.

Ivanek, Ferdo, *Terrestrial Digital Microwave Communications,* Artech House, Norwood, MA, 1989.

6.2
TELEVISION TRANSMITTERS

GERALD W. COLLINS, ROBERT J. PLONKA
HARRIS CORPORATION, QUINCY, IL

INTRODUCTION

Significant advances continue in TV transmitter technology. New technology and ideas have been introduced to continue to provide high quality TV signal transmission while improving reliability, reducing maintenance and lowering overall cost of ownership. These new technologies include solid-state high power amplifiers, improvements in UHF tube transmitters and digital transmission. Further, the FCC continues its policy of technical deregulation which allows more flexibility in transmitter design and system operation. This chapter discusses the relevant technology and provides the information needed for selection, installation, operation and maintenance of TV transmitters in today's environment.

TV transmitters may be considered to comprise two essential components, the *exciter* and the *RF power amplifier* (PA). The exciter provides the signal processing functions required to convert a baseband TV signal into a modulated RF signal on the assigned channel. These functions include baseband signal processing, modulation, precorrection and equalization, upconversion, bandlimiting and amplification to a relatively low power RF signal. While different methods are used for different signals, all functions must be performed whether the baseband signal is video, audio or digital data for digital television (DTV). Because the output of the exciter is a modulated RF signal, most commercially available exciters may be considered low power transmitters.

The PA provides the muscle to amplify the modulated RF signal to the desired level for transmission. Thus the PA technology to perform this function is key. Both solid-state and tube devices are available in commercial equipment. For VHF channels, solid-state devices predominate. For UHF, both solid-state and tube devices are used. Other key functions in the PA, common to all amplifier technologies include ac distribution, ac to dc power conversion, cooling and control.

TV transmitters are unique in that no other application requires such high levels of linear RF power generation while operating virtually uninterrupted. This has led to the development of specialized techniques to assure highly efficient and reliable operation. The need for high effiency has led to the near universal use of partially saturated Class AB final power amplifiers. This, in turn, has resulted in the development of precorrection and equalization techniques required to compensate for nonlinearities inherent in this class of operation. To achieve the levels of reliablility required, redundant system architectures that minimize single point failures are used.

EXCITERS FOR ANALOG TV

Exciter performance is the key to excellence in TV transmission. The exciter performs baseband video and audio processing, modulation, precorrection, equalization and upconversion functions. Some of these functions are common with exciters for digital transmission.

Visual Modulator

The *visual modulator* receives a video baseband signal, processes and converts it to a fully modulated vestigial sideband signal. Since intermediate frequency (IF) modulation is used, most of the signal processing occurs in the video stages and IF stages. The basic block diagram of an IF modulated transmitter is shown in Figure 6.2-1. For this discussion, the modulator is considered to include all blocks through the IF processor.

Video Processing

For transparent transmission of the incoming signal, it is important to optimize the incoming signal. It is often difficult to define where the station processing ends and the transmitter video processing begins. Typically some video processing is done external to the transmitter by a processing amplifier.

The main functions of the exciter video processing circuitry are:

- Obtain proper sync to video ratio
- Remove common mode signals
- Provide overall video level control
- DC restoration
- Prevent overmodulation
- Frequency response correction

Direct current restoration (clamping) is important because picture brightness information is contained in the dc component of the video signal. If ac coupling is used in the video circuitry, the dc level will tend to vary as the capacitors in the coupling circuits charge and discharge. Alternating currents coupling is convenient since differences in ac and dc ground can be allowed without introducing distortion. By clamping a particular level (such as sync tip or *back-porch*)

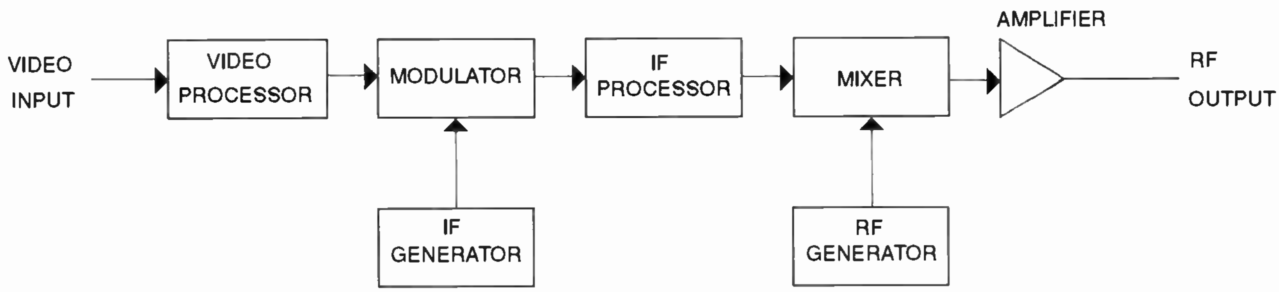


Figure 6.2-1. IF modulated transmitter.

during each line to a fixed voltage, the correct dc level is applied to the modulator.

Common mode signals such as noise or ac hum can be removed by using a differential input for the video input stage. It is desirable to use an RF choke at the video input to prevent rectification of ambient RF and remodulation of the main signal.

White peak limiting is used to prevent modulation from reaching 0% or to keep the carrier from being pinched off as shown in Figure 6.2-2. When an intercarrier TV receiver encounters a carrier that has been pinched off, it temporarily has no signal to receive and the automatic gain control (AGC) adjusts to maximum gain. Since there is nothing but noise, the noise is greatly amplified, transferred to the aural intercarrier and becomes audible as a buzz.

High-frequency peaking circuits are often used to compensate for long video signal transmission line runs in a transmitter plant.

In UHF exciters, the horizontal and vertical sync portions of the TV signal may be detected and sent to klystron pulsers.

Precorrection can be accomplished by changing the gain of a video amplifier over a portion of the total video signal. The threshold of the gain change is set at a certain level of the video signal. Differential phase can be corrected by splitting the video signal into two paths which are in quadrature at the color subcarrier. Phase changes can be obtained by changing the gain of only one path over a portion of the video waveform.

Modulator

In nearly all IF modulated transmitters, the modulator is a broadband, balanced diode mixer. It is configured for maximum rejection of the local oscillator signal and biased so as to provide excellent linearity, low noise and capability to achieve carrier cutoff. A schematic of a typical balanced mixer is shown in Figure 6.2-3. The video signal is dc offset to provide the proper modulation level and is used to control the attenuation of the diodes. Peak of sync corresponds to maximum IF envelope output and white corresponds to minimum IF output. The output signal of the modulator is a double sideband AM signal having the proper depth of modulation (12.5%).

Visual Group Delay Compensation

Group delay distortion appears as color smear and halo effects on edges. On a waveform monitor the effects may be seen using the 2T and modulated 12.5T contained in the composite test signal (see Figure 6.2-4). Pulse responses with unacceptable group delay may include exaggerated pre- or post-ringing on 2T and modulated 12.5T base line disturbance. Low-frequency group delay and amplitude errors are referred

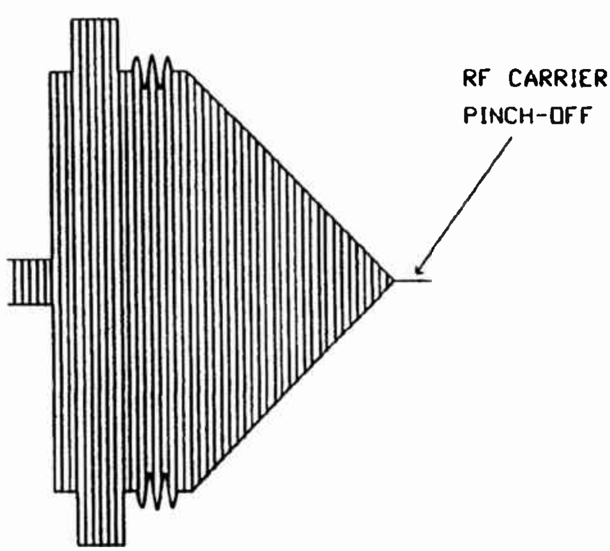


Figure 6.2-2. RF carrier pinch-off.

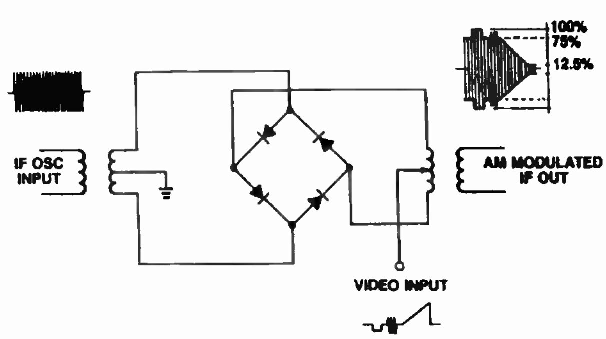


Figure 6.2-3. Balanced mixer.

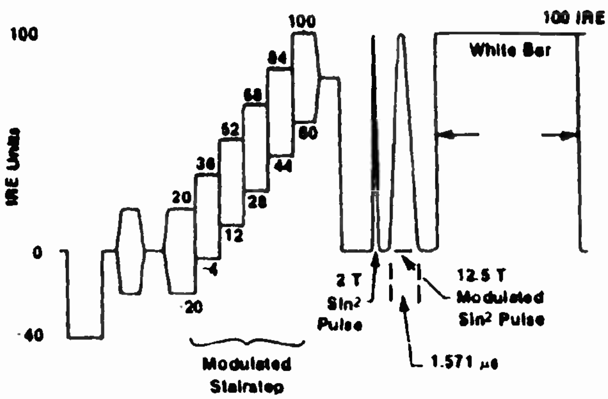


Figure 6.2-4. Composite test signal.

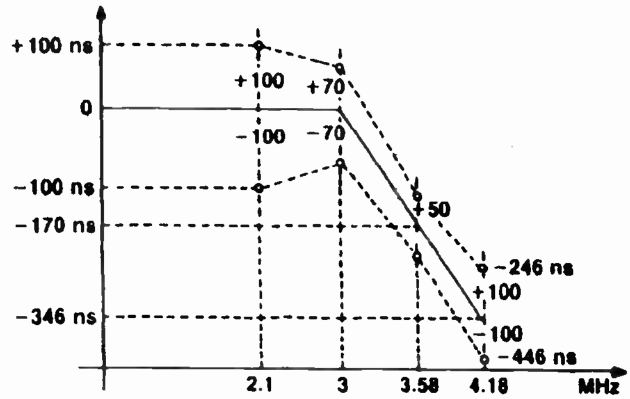


Figure 6.2-6. Predistorted group delay curve.

to as *short-time waveform distortions*. Figure 6.2-5 shows typical distortions of a 2T pulse.

Intermediate Frequency and Video Delay Compensators

The techniques used to accomplish group delay equalization at baseband or IF are similar in concept. Both active and passive equalizers may be employed. Intermediate frequency group delay correction is necessary to correct group delay errors below visual carrier while not affecting the signal above visual carrier. Above visual carrier, both IF and video correctors are effective. If the group delay error affects only the high-frequency side of the passband, equalization can be accomplished at video. This is the case for most visual-aural notch diplexers.

The FCC requires the TV transmitter to predistort group delay according to the curve shown in Figure 6.2-6. The purpose is to compensate for the delay caused by discrete LC aural notch filters in TV receivers. The notch filter in TV receivers is necessary to prevent the aural carrier from mixing with the detected video and chroma subcarrier signals and producing visible spurious beats.

Passive Group Delay Equalizer

A common form of a passive group delay equalizer is shown in Figure 6.2-7. It provides a flat frequency response and a nonlinear group delay which peaks at resonance. This circuit is referred to as a *passive all-pass network*. It can be understood by examination at frequencies well below and above resonance. At low

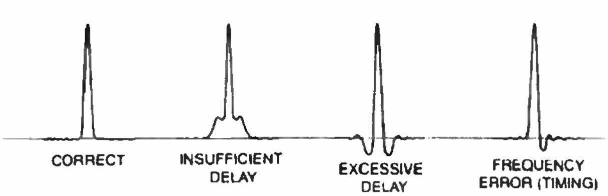


Figure 6.2-5. Typical distortion of a 2T pulse.

frequencies the network can be approximated by a low series inductive reactance due to L1 in which the output voltage leads in phase. At high frequencies well above resonance, the circuit can be approximated by a series capacitive reactance due to C1 in which the output voltage lags in phase. The output is constant across the band. The phase of the network is plotted in Figure 6.2-8. A plot of the group delay is shown in Figure 6.2-9.

Active Group Delay Equalizer

As in the passive type, this network has a constant amplitude versus frequency response and a nonlinear phase response. The action of an active *all-pass network* can be best seen using the simplified schematic of Figure 6.2-10. The phase of voltage (e_1) plotted as a function of frequency traces out a circle with maximum amplitude and zero phase at resonance. Voltage (e_2) has a constant amplitude and phase.

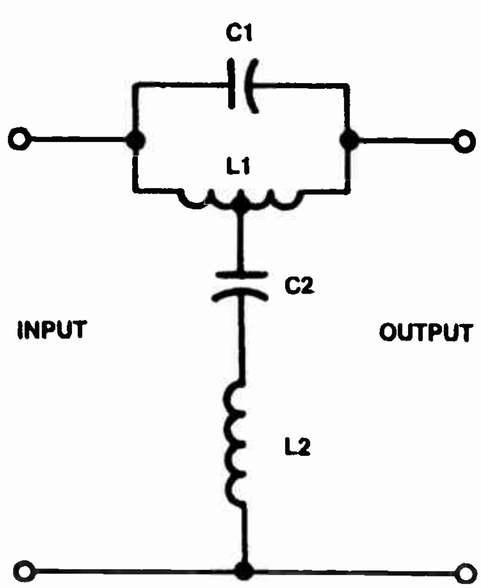


Figure 6.2-7. Passive all-pass network.

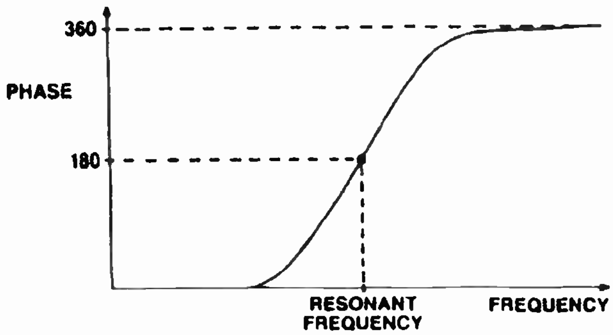


Figure 6.2-8. All-pass network phase response.

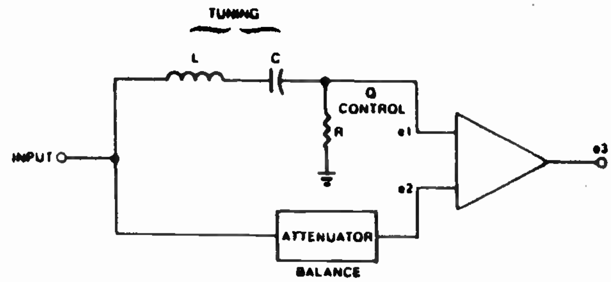


Figure 6.2-10. Active all-pass network.

When voltage e_1 is equal to twice (e_2), the output voltage of the summing amplifier, (e_3), has the characteristics of an ideal all-pass network. The output amplitude is constant and the resonator tuning determines the frequency of maximum group delay. The resonator "Q" determines the magnitude of delay. For the case where voltage e_1 is larger than twice e_2, the output has an amplitude peak at resonance. If e_1 is smaller, the output has a dip at resonance.

Vestigial Sideband Filter

The FCC requires that the radiated signal have a major portion of the lower sideband (vestigial sideband) suppressed. In addition, the upper sideband signal must be contained within 4.75 MHz of the visual carrier. With IF modulation the filtering is in the low-power stages. Transmitter manufacturers use solid-state filters using *surface acoustic wave* (SAW) technology to accomplish the filtering requirements.

The electrodes in a saw filter act like tapped delay lines and are designed to scale the amplitude and delay the signal (see Figure 6.2-11). The length, spacing and number of the electrodes determine the wave shaping properties of the filter. The output is selectively attenuated depending on the time delay and signal frequency relationship. This type of filter is called *transversal* because the attenuation is controlled by delay lines rather than resonators. The wavelength of an IF acoustic signal is approximately 0.003 in. The short wavelength allows the filter to be very compact. Because the transducers are on the surface, photographic masks can be used to accurately control their physical dimensions. The photographic mask lends itself to modern manufacturing techniques and insures a reproducible filter with a permanent amplitude versus frequency response and group delay.

A characteristic of a transversal filter is that group delay can be set independently of the amplitude. The independent nature of group delay and amplitude allows a manufacturer to provide almost any type of group delay curve across the passband, such as a constant delay or the TV receiver group delay equalization curve.

Although the FCC does not require attenuation of visual signals in the aural passband, video signal components can cause interference called visual-to-aural crosstalk. Reduction of visual-to-aural crosstalk is essential to the proper transmission of stereo sound and other subcarrier broadcast services. High resolution cameras and character generators can produce spectral components at 4.5 MHz and beyond. Without sufficient attenuation these signals can cause distortion, especially in aural subcarriers which are more sensitive to visual crosstalk than the main aural signal. SAW filters

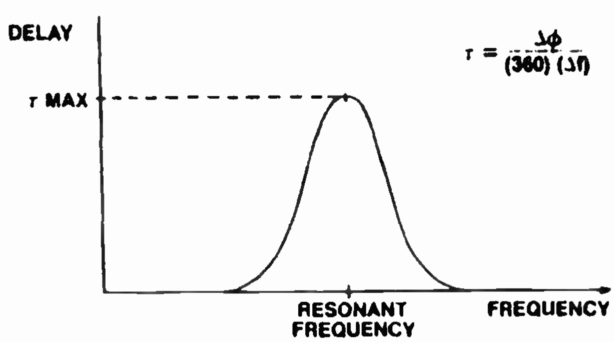


Figure 6.2-9. All-pass network group delay.

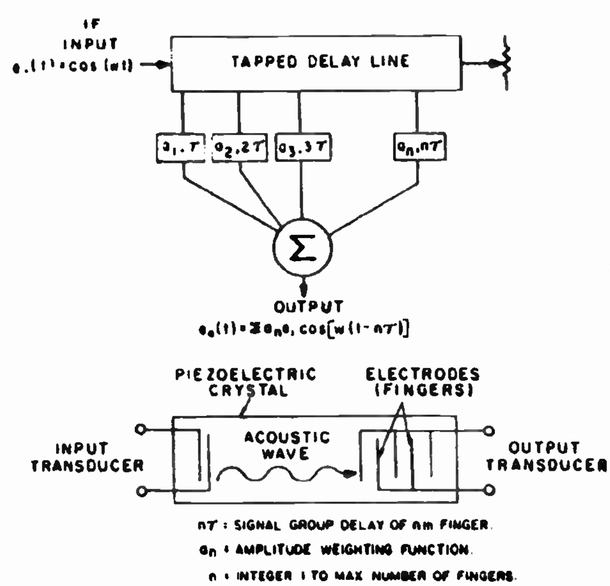


Figure 6.2-11. Saw filter.

thus are able to perform many functions simultaneously (bandwidth shaping, group delay equalization and visual to aural crosstalk reduction).

Distortion products can be created by SAW filters due to signal reflections (triple transit) and direct feedthrough. These distortions cause time echoes displaced either before or after the desired responses. In the frequency domain these echoes show up as ripples in the passband (see Figure 6.2-12). A given echo will contribute uniquely to the amplitude, phase and group delay characteristics by the addition of a ripple component in the passband. The amplitude of the ripple is proportional to the echo level and its period is proportional to the reciprocal of the SAW time delay.

SAW group delay varies sinusoidally with the same period as the passband ripple. Magnitude, however, is a function of echo level and is inversely proportional to time delay. It is, therefore, not a sure test of signal distortion. The group delay ripple for long delayed echoes will give peak-to-peak values which have no correlation with conventional signal distortion estimates. For that reason, fast peak-to-peak group delay errors that occur at greater frequency than the reciprocal of the filter time delay can usually be ignored.

Intermediate Frequency Linearity Precorrection

An alternative to video correction is to provide the correction at IF. There are advantages to correcting

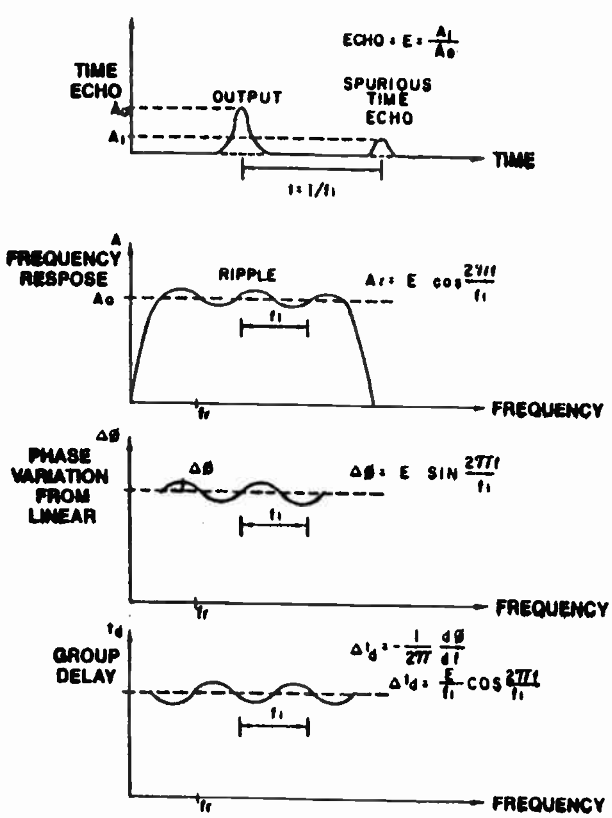


Figure 6.2-12. Response, phase, and delay ripples caused by time echoes.

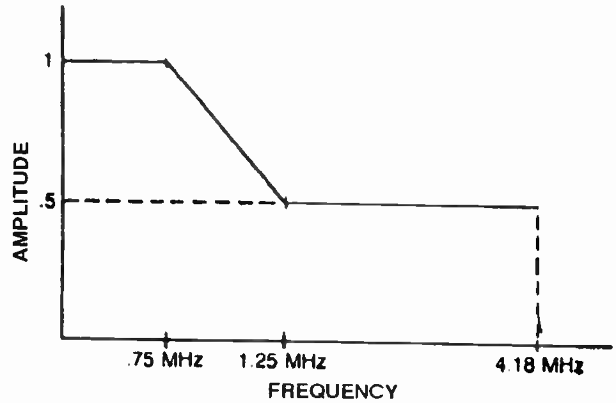


Figure 6.2-13. Frequency response of ideal diode detector.

at IF. Since most distortions are caused in the high power RF amplifiers after vestigial sideband filtering, a corrector placed after the vestigial sideband filter can more accurately predistort the modulated signal. Any precorrection spectra generated at IF after the VSB filter will produce energy components which can cancel intermodulation products generated in the final amplifier stage. This is particularly important in the cases of pulsed klystron transmitters or common amplification.

Correction of distortion at IF is particularly helpful for chroma distortions. Chroma spectra at 3.58 MHz has only single sideband information and thus less energy than equivalent luminance signals. An ideal diode detector frequency response of the NTSC modulation signal is plotted in Figure 6.2-13. It is seen that beginning at 0.75 MHz the video begins to fall to 6 dB. For video signals lower than 0.75 MHz, the RF spectrum is double sideband and has twice the peak RF voltage.

Intermodulation products are caused in high-power amplifiers by a nonlinear transfer function. As the power output increases towards saturation, amplitude compression and phase lag occurs. The nonlinear transfer function gives rise to mixing products which occur at sum and difference frequencies around the visual carrier. This process creates the frequency spectra of the lower sideband, commonly referred to as *lower sideband reinsertion*.

Intermediate frequency correctors for amplitude distortion are usually similar in concept to video differential gain correctors. Linearity correctors generally use diodes which are set to conduct at a specific level of the IF modulated signal. When the diodes conduct, the gain or attenuation is reduced as needed.

An example of a basic gain expansion circuit is shown in Figure 6.2-14. The signal is normally attenuated a fixed amount by using a resistive L-pad. The diodes are normally reverse biased by equal dc voltages. Reducing the dc voltage permits the diodes to conduct on the signal peaks. This inserts additional resistance in parallel with the series arm of the L-pad attenuator thereby decreasing the attenuation. Varying

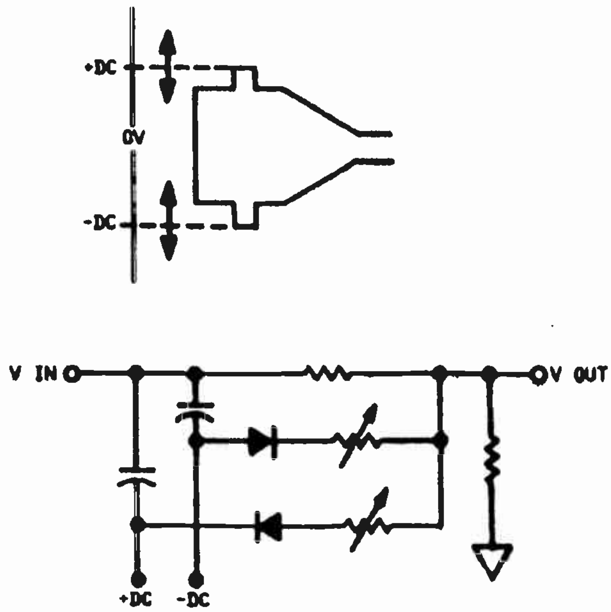


Figure 6.2-14. Basic gain expansion circuit.

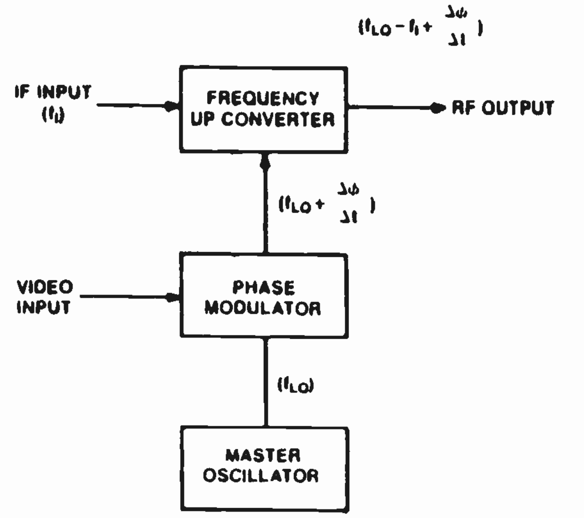


Figure 6.2-15. Master oscillator phase modulator block diagram.

the resistance in series with the diodes provides for variable gain expansion.

Incidental Carrier Phase Modulation (ICPM)

Phase distortions in high-power amplifiers produce ICPM or spectral components in quadrature with the modulation signal. Fast video amplitude changes such as a step or pulse will cause larger incidental phase spectral components than slow changes. Receivers make this condition worse by attenuating the lower sidebands below 0.75 MHz. The receiver then responds to the extra sidebands created by the phase modulation as if they were amplitude modulated single sidebands, producing spikes. The faster the rise time of the signal, the more high-frequency energy is present, resulting in edge distortions in the displayed picture.

The picture impairment is similar to simultaneous group delay and differential phase errors in that edges are less sharp and color hue changes with brightness. On a waveform monitor, overshoots are visible as trailing edges and as rounding on leading edges. These overshoots vary in severity depending on how close to saturation the power amplifier is driven.

Audio impairment is produced by ICPM in receivers employing intercarrier conversion. Intercarrier receivers use an AM or synchronous detector to produce a 4.5 MHz aural IF from the composite video IF. Any phase modulation present on the visual carrier is then transferred to the aural intercarrier. In monaural baseband audio, the increasing amplitude versus frequency effect of ICPM is nullified by deemphasis to some degree. With multichannel sound, however, there is no deemphasis applied to the baseband stereo signal, and thus the distortion is more pronounced at the stereo subchannel and pilot frequencies. To counteract the effects of ICPM and other noise sources on the stereo subchannel, audio companding is employed. Although the audio companding process can reduce some of the effects of ICPM, correction is essential in delivering clear, low-noise audio to intercarrier receivers.

There is no defined level of ICPM for a given stereo performance level since the signal to buzz ratio is highly dependent on the picture spectral components. (Refer to the *EIA Recommended Practices* for current recommendations on ICPM limits.)

ICPM precorrectors can be grouped into two types: those using a phase modulator and those operating on the signal directly. The phase modulator uses video to modulate the IF or master oscillator with a phase characteristic opposite that of the nonlinear amplifier. A phase modulator can also operate on the IF signal directly using a video signal to set the amount of modulation. A block diagram of a master oscillator phase modulator is shown in Figure 6.2-15.

ICPM precorrectors operating directly on the IF signal can be implemented several ways. Direct precorrection at IF is similar in concept to baseband differential phase precorrection (see Figure 6.2-16). In both cases, the visual signal is split into two paths which are in phase quadrature. In the IF corrector, the entire

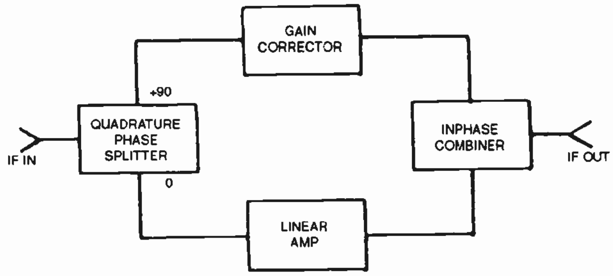


Figure 6.2-16. Direct IF ICPM corrector.

channel of frequencies is in quadrature, whereas in the video precorrector only the chroma band is in quadrature. One method of implementation is to modify the quadrature signal gain function with level dependent diode expansion or compression circuits. This can be done using the same techniques as in the linearity corrector.

The vector diagram in Figure 6.2-17 illustrates nonlinearities and the operation of the ICPM and linearity correctors. The desired output signal is represented by the left vector. Uncorrected the transmitter output signal is phase shifted and compressed. To compensate, the signal in the exciter is precorrected by an amplitude expansion and a correction in quadrature. When the resultant signal is amplified, the output signal is the desired TV signal.

Aural Modulator

In its most basic form, the aural modulator consists of some audio processing and an FM modulated IF oscillator.

Audio Processing Circuits

To ensure that the transmitter is not the limiting factor in audio (monaural and stereo) reproduction, it is desired that the transmitter add as little distortion to the incoming signal as possible.

Baseband audio of the BTSC Multichannel Television Sound (MTS) System (mono, stereo, second audio program (SAP) and professional channels (PRO)) include frequency components to 105 kHz. Emphasis must be placed on phase linearity, low distortion, reduction of any amplitude ripples and roll-off over the stereo passband to achieve good stereo separation and minimum crosstalk between the main stereo and the SAP channels.

While unbalanced coaxial inputs are used for the MTS input to the aural exciter, it is necessary to use some form of common mode rejection to reduce the possibility of hum and noise from getting directly into the audio stages.

All errors in phase linearity and amplitude response within the audio circuitry contribute to stereo separation degradation. As a general rule, amplitude roll-off should be less than 0.1 dB and departure from phase linearity should be less than 1° for good quality stereo.

IF Modulation

With the MTS system, intermodulation (IMD) and harmonic distortion products lie in the stereo channel or the SAP channel and will degrade stereo separation or crosstalk into the SAP channel. In addition, IMD products generated in the stereo channel may lie in the mono or SAP channel. Modulated oscillator linearity requirements include flat modulation sensitivity versus frequency up to 47 kHz minimum, and typically, out to 120 kHz.

As the MTS signal is upconverted to the aural RF carrier frequency, the residual FM of the local oscillator signal becomes a determining factor. The level of FM produced by the local oscillator should be 10 dB lower than the modulated oscillator. Synthesized sources should be tested for spurious frequencies which may show as FM noise. To ensure low-power stages do not contribute any group delay or amplitude roll-off, wideband amplifiers should be used.

Compensation of Aural Passband for Optimum Stereo Performance

Group delay equalization for the aural transmitter IF section of the aural modulator effectively corrects the adverse group delay in the diplexer. The end result is improved TV stereo separation. Equalization of the FM bandpass allows the use of lower cost, single-cavity notch diplexers. Stagger-tuned dual-cavity notch diplexers have been used to provide the broad bandwidth desired for good stereo separation and neg-

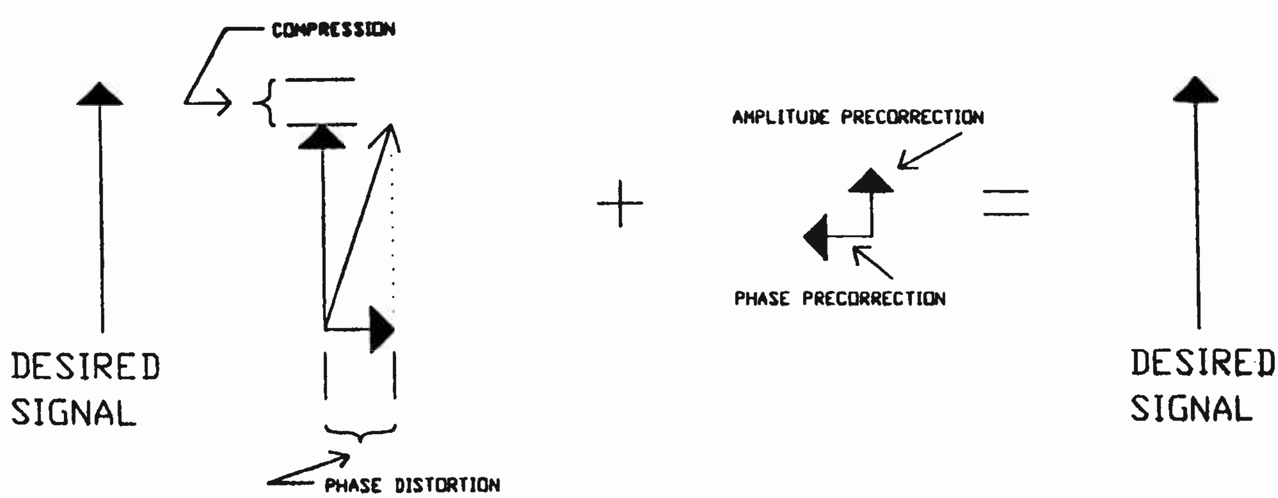


Figure 6.2-17. Vector representation of precorrection.

ligible crosstalk between the different MTS components. However, dual-cavity notch diplexers introduce more group delay in the visual path and are more expensive than single-cavity diplexers.

The notch diplexer is a passive device and under first consideration it may seem strange that it can introduce distortion and stereo separation errors. The basic problem is that the FM stereo signal is sensitive to the notch diplexer group delay and amplitude response over the occupied bandwidth of the FM signal.

The group delay and amplitude response of a single-cavity diplexer is shown in Figure 6.2-18. It is a typical notch diplexer optimized for minimum aural reject power. The bandpass is somewhat narrow and the group delay is steep. Fortunately, the response curves show a high degree of symmetry which makes equalization possible.

The group delay equalization concept as applied here is essentially a *feed-forward scheme.* Feed-forward is used to describe the technique of generating equalization early in the RF lineup for the purpose of correcting distortions occurring downstream in the system. Feed-forward is operated open loop (without feedback) and is manually adjusted for optimum operation.

Aural IF Group Delay Equalization

The equalization technique of adjusting the circuit to produce an inverse curve of that of the diplexer is the same as that used on the visual signal. There is, however, a significant difference between group delay correction on a FM system versus an AM system. The aural transmitter must incorporate group delay correction at RF or IF to compensate for group delay distortions occurring in the diplexer. *Group delay correction before the FM modulation process is not effective for equalizing group delay occurring at RF.* The reason for this is that the occupied bandwidth of an FM signal increases or decreases as a function of baseband signal level. Figure 6.2-19 shows the occupied bandwidth when the carrier is deviated 25 kHz compared to a typical notch diplexer group delay curve. When the baseband signal level is increased, the FM deviation increases and a number of additional significant sidebands are generated beyond the acceptable group delay curvature region. The result is distortion in the demodulated FM signal.

In addition, as the baseband high-frequency content increases (a stereo signal) significant sidebands extend even further increasing the demodulated distortion. Figure 6.2-20 shows the spectral content of a TV stereo signal with notch diplexer group delay curves overlayed. The equalizer location in the aural transmitter system is shown in Figure 6.2-21.

The circuit configuration is shown functionally in Figure 6.2-22. The measured amplitude and group delay is shown in Figure 6.2-23. An ideal precorrection circuit without any circuit losses would have a flat

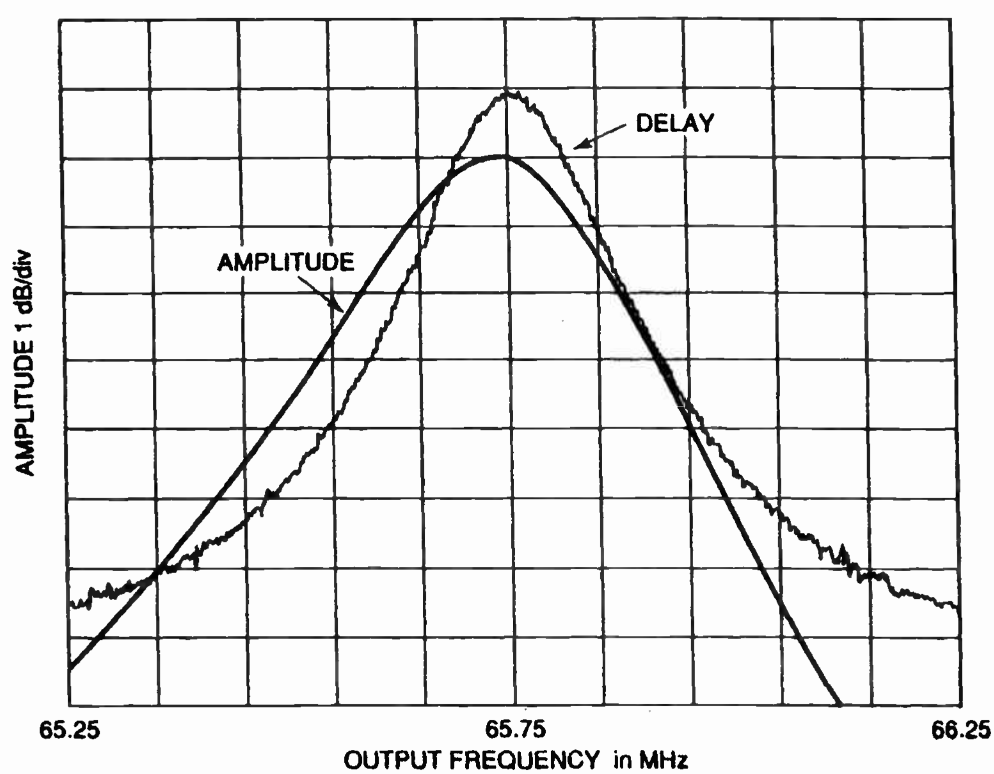


Figure 6.2-18. Typical single cavity diplexer amplitude and group delay.

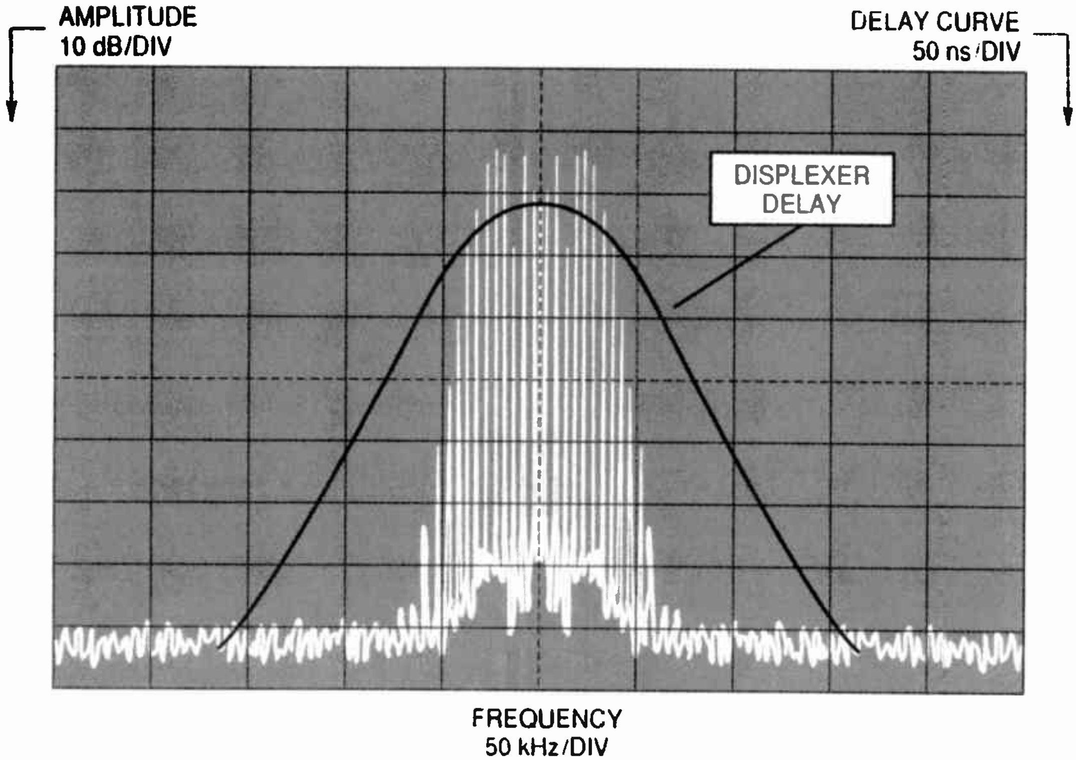


Figure 6.2-19. Occupied bandwidth with 25 kHz deviation with overlayed notch diplexer bandwidth.

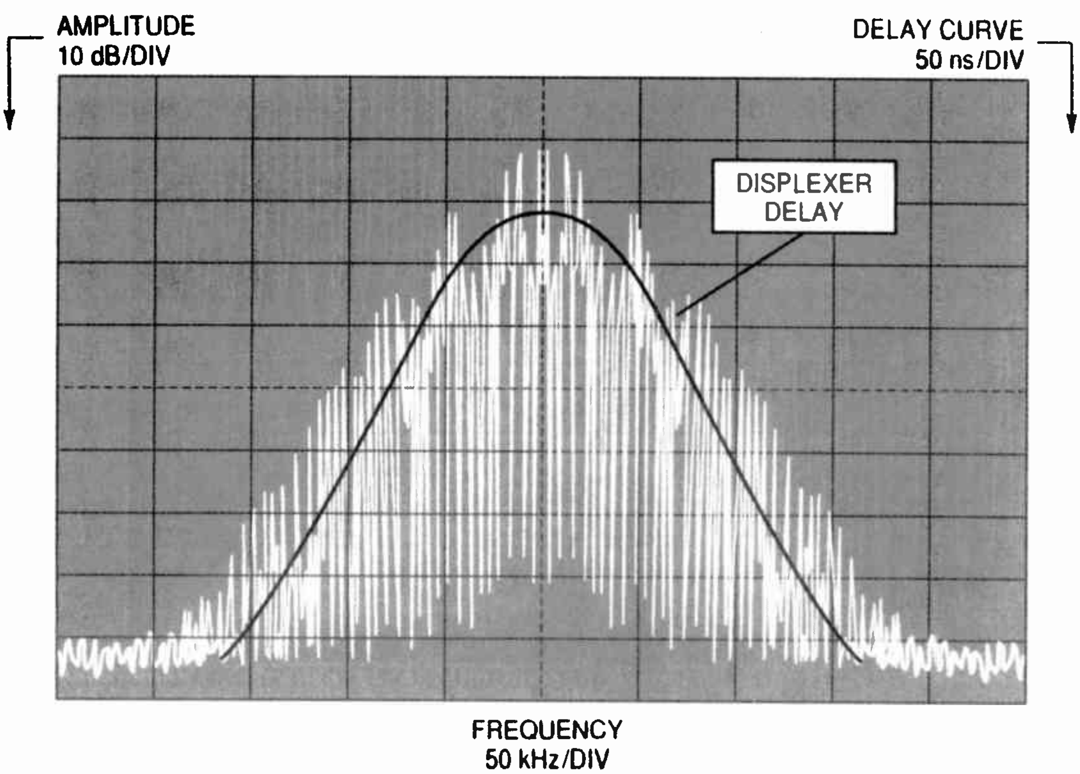


Figure 6.2-20. Occupied bandwidth of TV stereo signal with 55 kHz deviation and overlayed notch diplexer bandwidth.

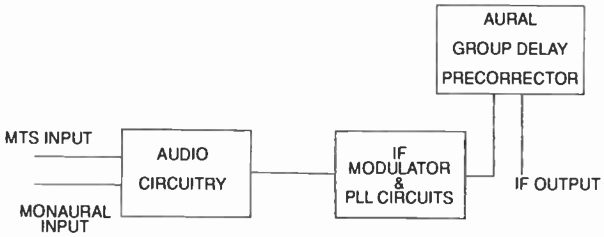


Figure 6.2-21. Aural indicator exciter with group delay precorrector.

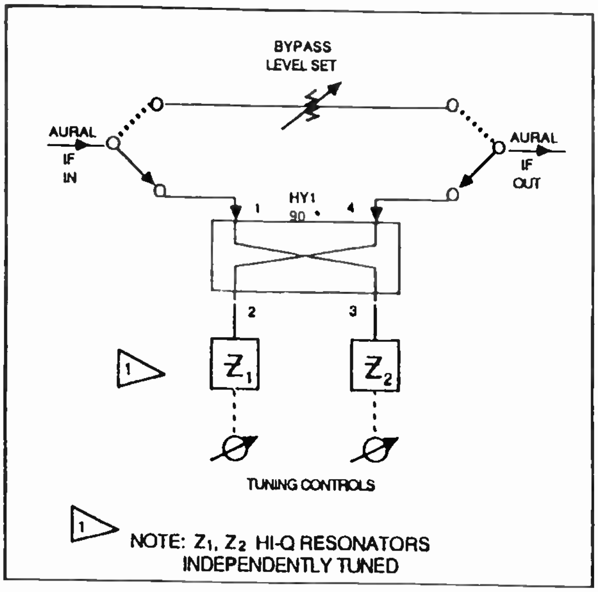


Figure 6.2-22. Aural group delay corrector functional block diagram.

response without a dip. The response dip is very useful, however, because it also provides a first-order correction to the notch diplexer amplitude response.

Figure 6.2-24 shows the system equalization when the delay corrector is switched in and switched out. A significant amount of equalization is achieved over the occupied bandwidth of a stereo signal.

The effect on stereo separation is shown in Figure 6.2-25. More than 10 dB stereo separation improvement can be obtained over midband audio frequencies.

Common Amplification

Traditionally, TV transmitters have used separate RF amplifier chains for the visual and aural signal paths. Assuming a single antenna and feed line, this requires a high level diplexer to combine the visual and aural signals prior to transmission. With the introduction of inductive output tubes as final amplifiers, it has become popular to combine the visual and aural signals in the exciter and amplify them together in the following stages.

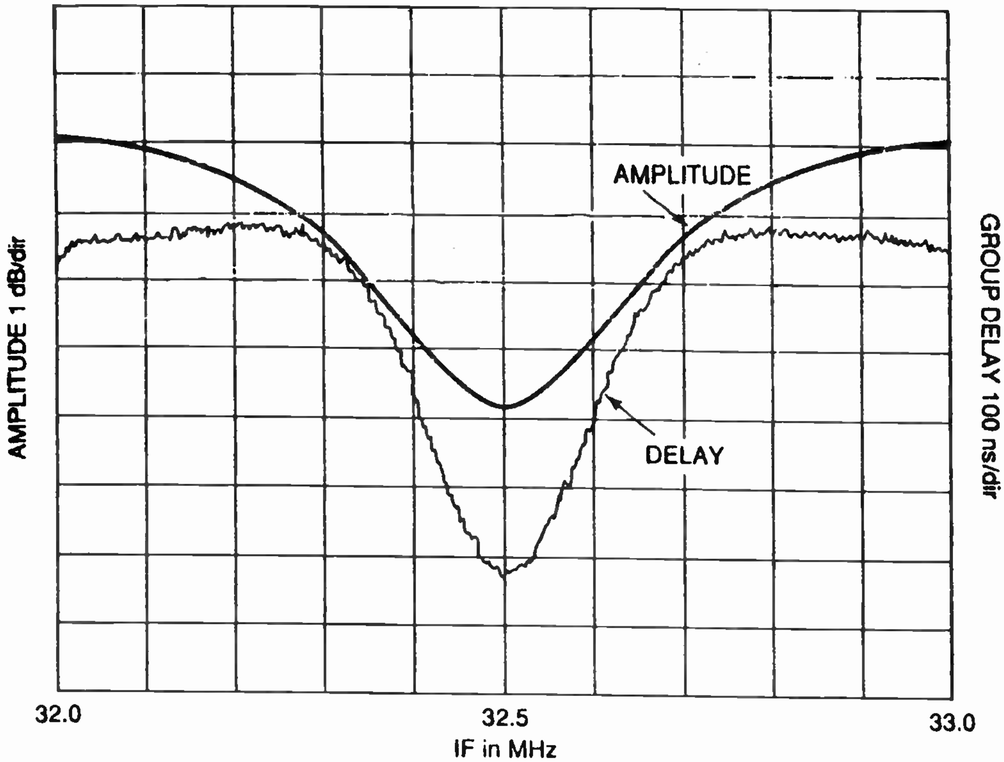


Figure 6.2-23. Aural group delay corrector measured results.

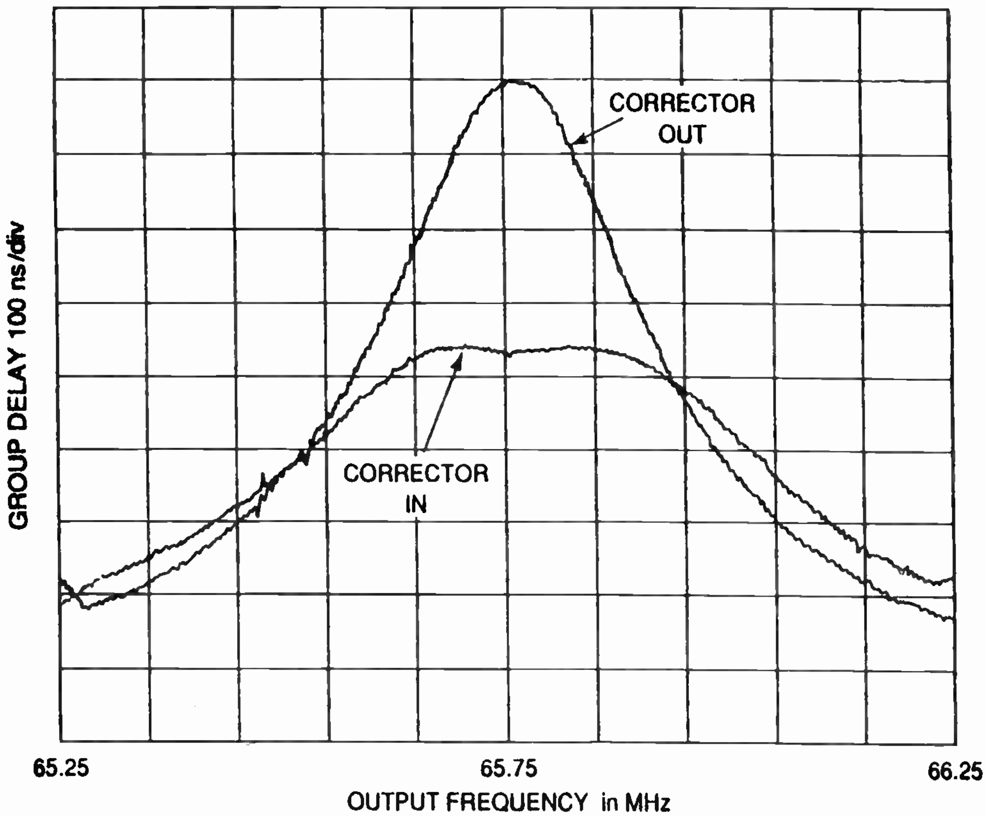


Figure 6.2-24. Overall transmitter delay with aural delay corrector switched in and out.

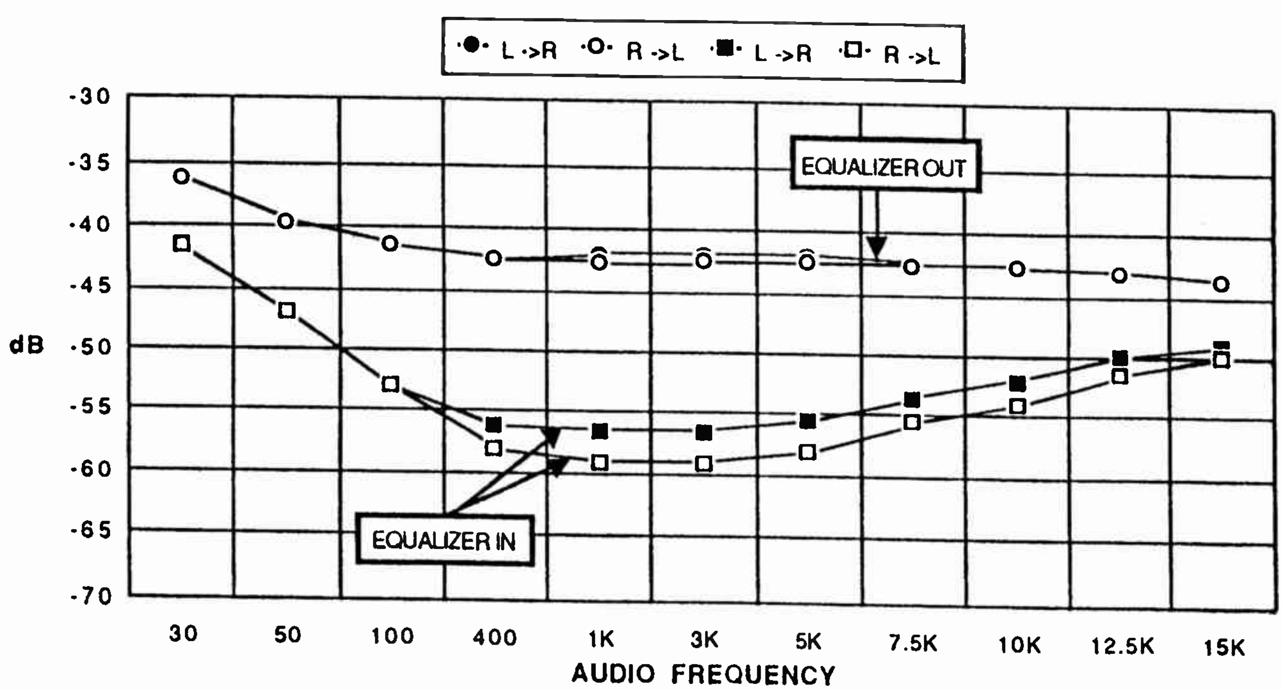


Figure 6.2-25. TV stereo separation with and without group delay equalizer.

Even though very linear amplifiers are used, some residual nonlinearity remains. This has the effect of mixing the visual, color and aural signals to produce in-band as well as the out-of-band IMD and cross modulation products. The out-of-band products sufficiently removed from the channel of operation are attenuated by the high level filter. However, in-band products can be removed only by making the transmitter sufficiently linear to reduce these products to adequate levels. This requires the use of highly effective correction circuits in the IF and/or RF signal paths. For example, the IMD at ±920 MHz can be precorrected by low level IF circuitry.

Up Conversion

Transmitters employing IF modulation generate the following frequencies: visual IF, aural IF, and master oscillator signal(s) for translating visual and aural IF to the final carrier frequencies. These oscillators have been implemented with either digital synthesizer techniques or crystal oscillators. An advantage of the synthesizer is that only one crystal is needed at a single standard frequency for all TV channels. The crystal oscillator approach, however, may involve simpler circuitry.

The two commonly used IF frequencies are 37 MHz and 45.75 MHz. There are many reasons for selecting one IF frequency or the other. One advantage of 37 MHz is that the temperature drift sensitivity of most IF components such as the SAW filter, is related directly to carrier frequency. Thus the lower IF has a 12% less drift sensitivity than components at 45.75 MHz. The second harmonic of 37 MHz falls in between channels 4 and 5 so as not to cause interference. On the other hand, 45.75 MHz is a common demodulator IF which can be useful for IF troubleshooting. Temperature drift may be minimized at either IF by maintaining the SAW filter at a stable temperature.

The important performance characteristics of an oscillator are phase noise, frequency stability over time and temperature extremes and level of microphonics.

In replacing a crystal, it is important to follow the recommendations of the oscillator manufacturer to ensure proper operation. Synthesizer performance should be properly maintained to prevent inadvertent phase noise and spurious frequency generation.

Offset Frequency Control

The limited number of available channels for TV broadcasting makes it necessary to assign the same carrier frequencies to many stations. To avoid interference between stations operating on the same frequency (co-channel interference), geographical separation and radiated powers are carefully selected.

Co-channel interference between television stations appears to viewers as a horizontal pattern of alternating light and dark bars on the viewing screen—very much like the shadows cast by venetian blinds. It has been demonstrated that the visibility of these bars varies cyclically as a function of the difference in frequency of the interfering carriers (see Figure 6.2-26). The

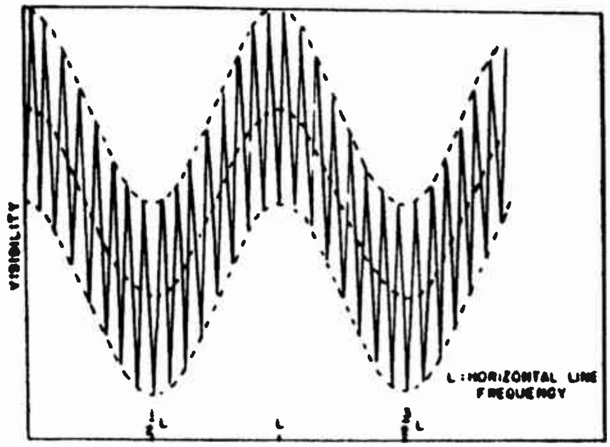


Figure 6.2-26. Co-channel interference.

interference is least visible when the carriers are offset by odd multiples of one-half the line frequency (15734 kHz). Fine grain maxima and minima occur when the frequency offset is an integer multiple of the frame frequency (29.97 Hz).

Ideally, stations would be offset by odd multiples of one-half the line frequency to provide minimum interference visibility. However, a third station in the same area would be offset from one of the other stations by an even multiple of the line frequency. Hence, maximum visibility of the interference would occur. Therefore, 10 kHz offsets currently are used in the United States to provide approximately equal reduction of the interference patterns for any number of stations in geographical proximity (see Figure 6.2-27).

Precise Frequency Offset

Although it is not practical to utilize the gross minima occurring at odd multiples of one-half the line frequency, use of the fine grain minima, occurring at even multiples of the frame frequency, is very advantageous in reducing the visibility of co-channel interference.

The nearest even multiple (334th) of the frame frequency to the 10 kHz offset is 10010 Hz. In a three station arrangement, one station will have zero offset and the other two stations will be offset by ±10010 Hz. Changes in the frequency differences of 5 Hz have a negligible effect on the reduction of the interference visibility.

To maintain the precision offset within 5 Hz requires maintaining each visual carrier frequency within 2 or 3 Hz. Maintaining a television transmitter to such tight frequency tolerances requires an extremely stable frequency source. (Note that 3 Hz stability is required in the case of an adjacent DTV signal.)

In exciters using two independent oscillators (one for IF and one for the local oscillator), the visual carrier signal may be derived from mixing the oscillators together and comparing that signal to the reference signal in the comparator. The resultant error signal can be

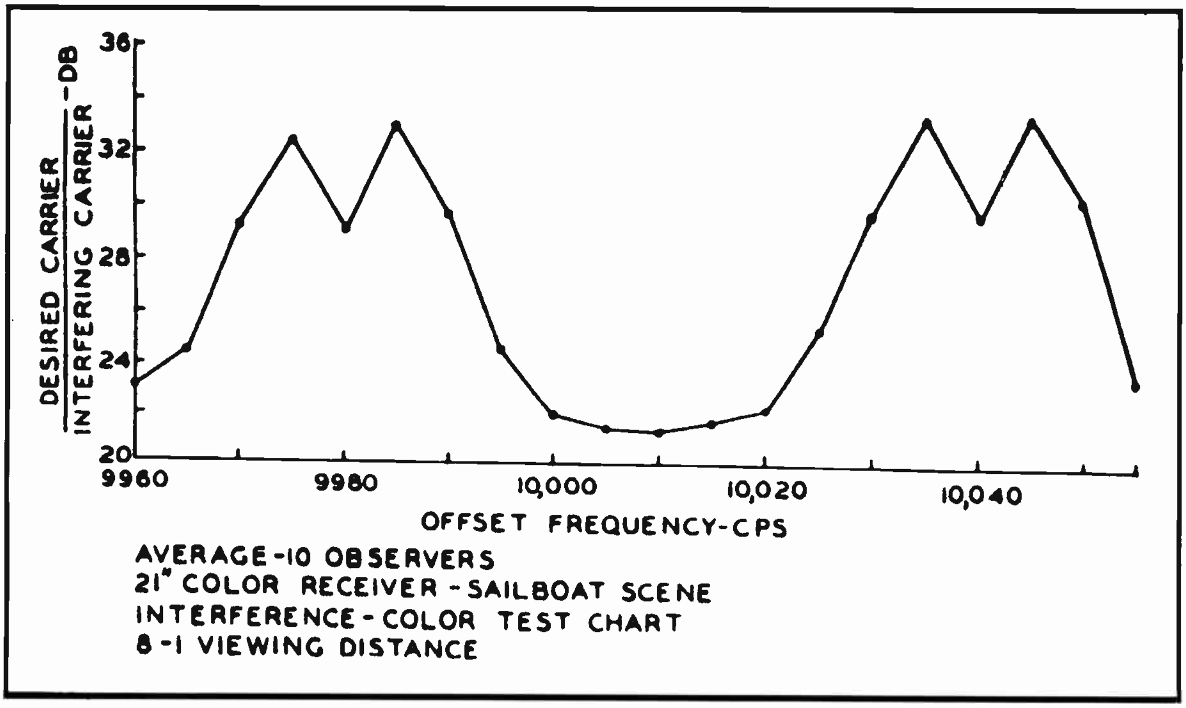


Figure 6.2-27. 10 kHz offset pattern.

used to adjust one of the oscillators, preferably the local oscillator. For exciters using synthesizers, the synthesizer reference oscillator is compared to the precise frequency standard in the phase detector and the resultant error voltage is used to adjust the synthesizer oscillator.

By phase locking the visual carrier to a stable reference oscillator, the master oscillator acquires the stability of the reference source. Sources which use an internal WWV receiver/comparator to self-correct or an atomic frequency standard can easily provide the stable reference source required. These reference sources allow a transmitter to be maintained within a few hertz of a desired frequency indefinitely.

RF Amplifiers

The last active stage in a typical exciter is the RF amplifier. For visual signals in separate amplification, combined aural and visual signals in common amplification and for digital signals, it is important that this amplifer have a linear transfer characteristic in amplitude and phase and flat, symmetric frequency response and minimum group delay variation across the modulation passband. For visual only signals, the required bandwidth is at least 4.5 MHz. For common amplification and digital signals at least 6 MHz bandwidth is required.

For optimum stereo performance a nonlinear amplifier with flat response and group delay across the modulation pass band is required. Since frequency modulation and demodulation is a nonlinear process these is not a one to one correspondance between RF amplitude/phase response and baseband stereo separation and crosstalk. A 3 dB bandwidth of 1.5 MHz provides excellent stereo and SAP performance.

EXCITERS FOR DTV

With the introduction of digital transmission technology, the broadcast engineer must be familiar with the equipment required to broadcast high speed digital signals as well as the more familiar analog video and audio signals. There are many aspects of exciter design that are common to analog and digital transmission. This is especially true of the upconversion and IF precorrection and equalization sections. Baseband processing and modulation are the portions of the exciter most affected by digital transmission.

The modulation used in the ATSC terrestrial broadcasting standard is 8-level vestigial sideband (8 VSB). Figure 6.2-28 shows the conceptual implementation of 8 VSB. Data are taken 3 bits at a time and converted to 8-level pulses which are applied to a baseband filter. The filter is offset in frequency by one-fourth the symbol rate to produce a baseband version of 8 VSB. There are many ways to implement the 8 VSB modulator, but the objective of all approaches is to produce the signal as in the conceptual representation.

Modulator

Figure 6.2-29 shows the functional block diagram of a typical 8 VSB modulator. Data formatted and

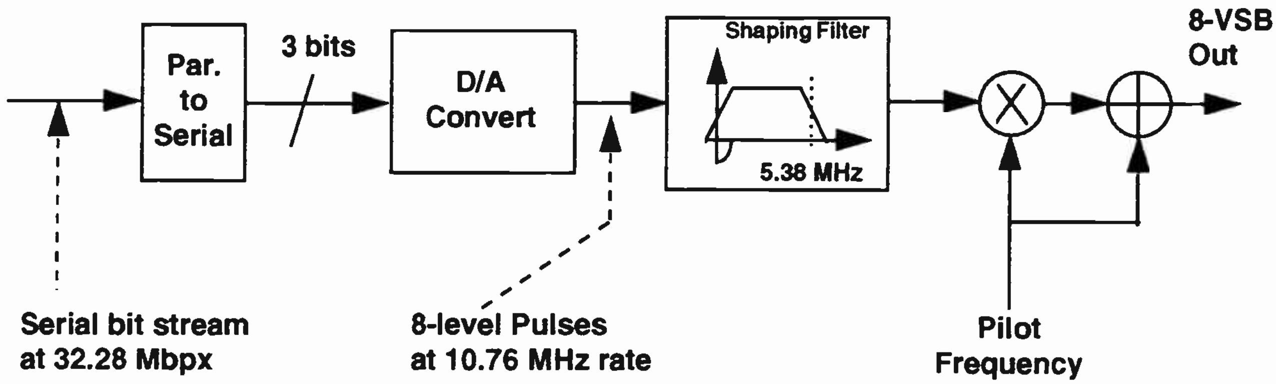


Figure 6.2-28. 8 VSB Conceptual block diagram.

encoded in accordance with ATSC specifications emerges from the transport-to-transmisson layer and is presented 3 bits at a time at the 10.76 MHz symbol rate to the baseband digital Nyquist filter. After digital filtering, the data is in the form of a 10.76 MHz digital IF. It is then converted to an analog IF signal using a high speed digital-to-analog converter (DAC). The analog IF output is filtered by a x/sinx correction filter, the purpose of which is to compensate for the sinx/x response inherent in the square pulses emerging from the DAC. A low-pass filter is also required to attenuate the image frequency before the first IF is upconverted to the second IF at 45 MHz. From the second IF the 6 MHz bandwidth 8 VSB signal is upconverted to carrier.

Signal Envelope of 8 VSB

Figure 6.2-30 shows the typical signal envelope of 8 VSB modulation. Note the great variability in the envelope peaks. This time-varying peaking is the source of the large peak-to-average ratio. The variability in the envelope gives rise to requirement of very linear power amplification. Figure 6.2-31 shows the spectrum measured from a typical modulator. The spike in the spectrum on the left side indicates the presence of the pilot carrier.

Hardware Design

The 8 VSB modulator consists of three major functions: the transport-to-transmission layer converter or channel coding, the Nyquist filter and IF conversion (see Figure 6.2-32).

Transport/Transmission Layer Conversion

The ATSC has defined the data transport layer to be a 19.389 Mb/s data stream. The data is formatted into 8 bit bytes, and 188 bytes per frame, each frame consists of 187 bytes of data plus one byte reserved as a sync byte. The resulting payload is the 19.389 Mb/s data stream.

Channel coding protects the data from noise and interference by adding redundancy prior to transmission. This redundancy is decoded by the receiver to correct for errors introduced by transmission. The modulator performs the channel coding by implementing frame synchronization, data randomization, Reed-Solomon (R-S) encoding, data interleaving, trellis encoding and field and frame sync insertion (see Figure 6.2-33). The purpose of the randomizer is to disperse the payload data energy in pseudo-random fashion to assure that the spectral energy is constant over the channel bandwidth. This is necessary to assure optimum transmission and reception. The next functions are designed to protect the data from transmission errors. This is a two stage process. R-S coding is used in the first stage. This is a forward error correction code suitable for correcting block errors due to bursts of noise and interference. The code selected is capable of correcting up to 10 byte errors per block. This is accomplished by adding 20 parity bytes to each 187 byte block of the payload. Interleaving of the data provides additional error protection. This convolutional byte interleaver spreads the data bytes from each R-S block over time so that a long burst of noise is required to overrun the R-S error protection. Trellis or convolutional coding is used as the second stage of

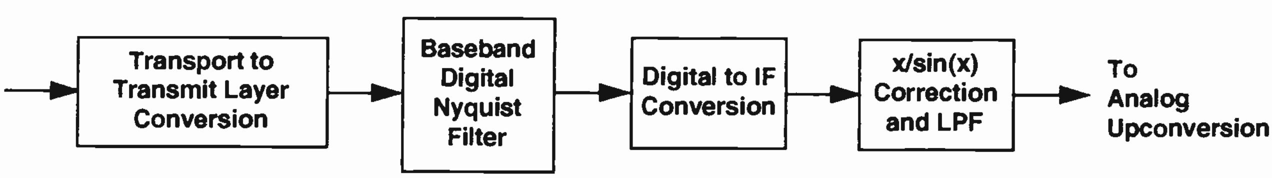


Figure 6.2-29. Typical DTV 8 VSB modulator.

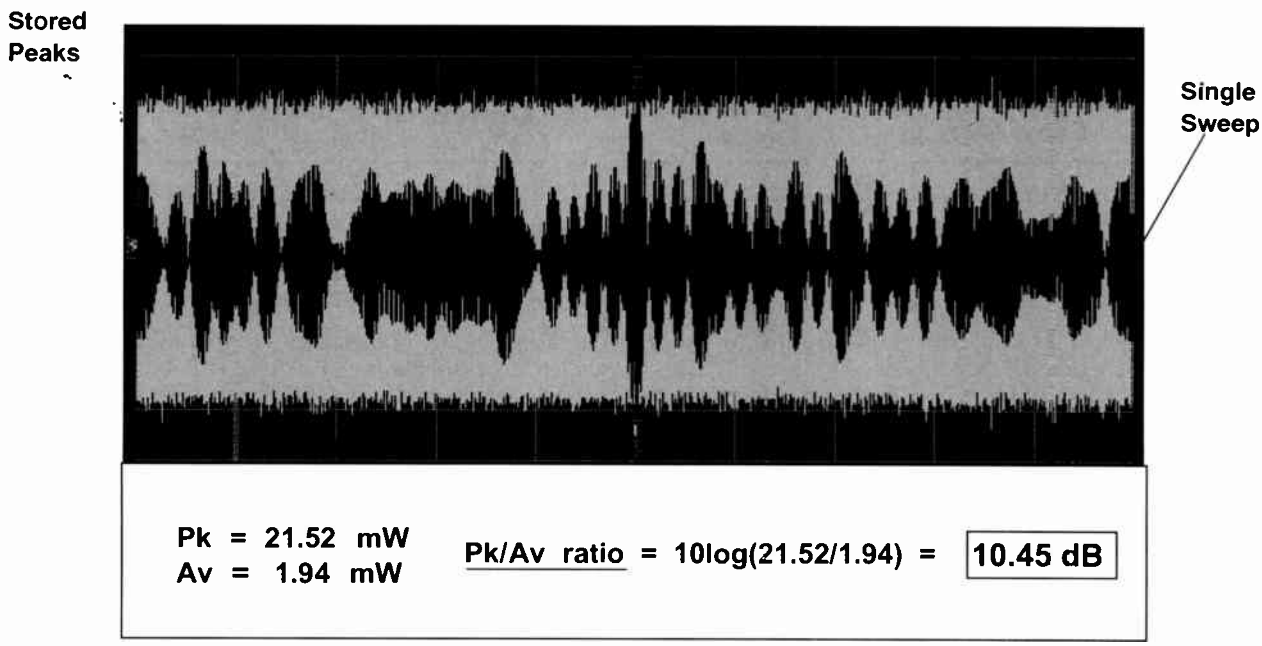


Figure 30
DTV RF waveform

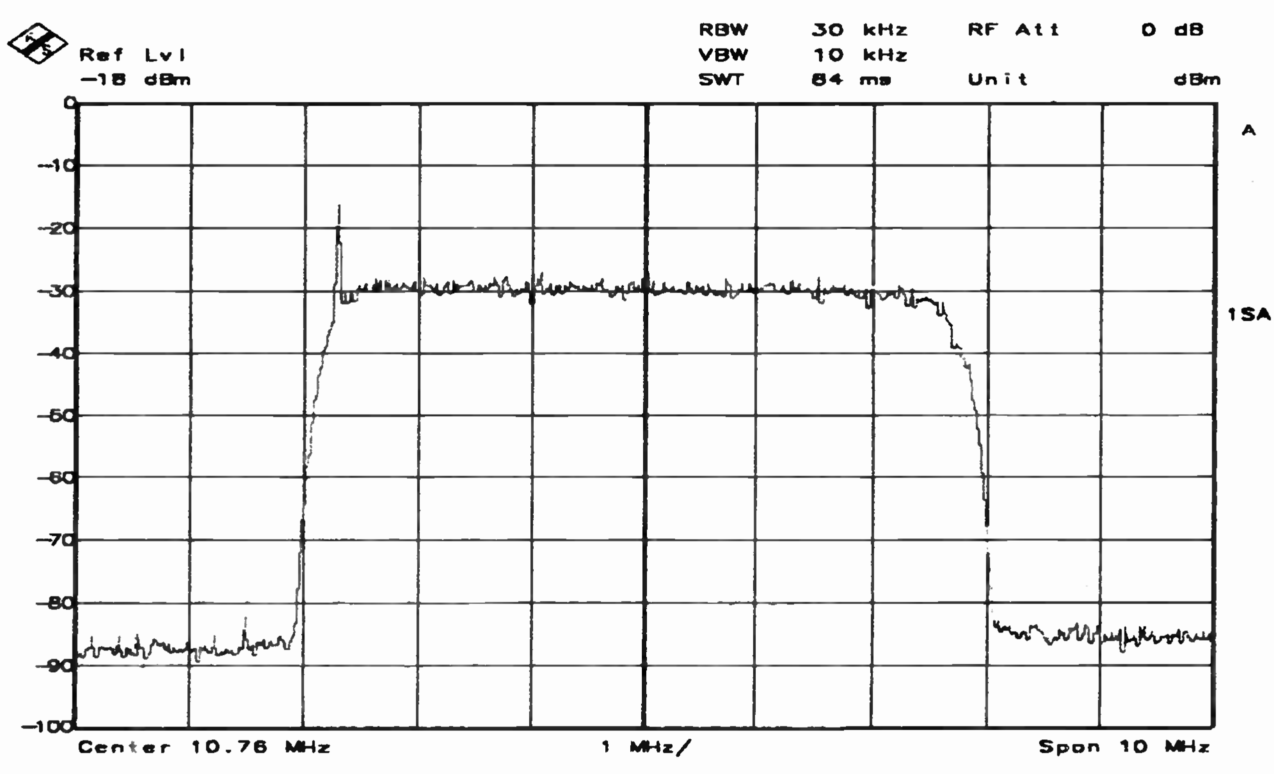


Figure 6.2-31. Measured spectrum.

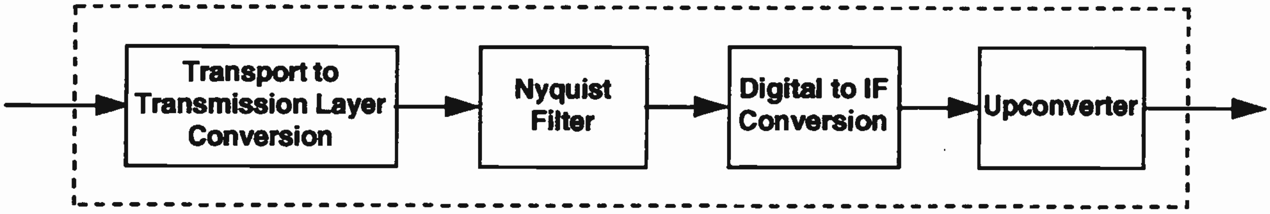


Figure 6.2-32. DTV exciter.

error correction. This is used to used to improve the S/N threshold against white noise. A 2/3 rate code with one unencoded bit is used. The signaling waveform used with the trellis code is an 8 level (3 bit) one dimensional constellation. Trellis code interleaving is also used. Following these encoding processes, field/frame and data segment syncs are inserted by means of a multiplexer.

Nyquist Filter

The *Nyquist Filter* is a root-raised cosine filter with ideal alpha factor of 0.1152. This filter, if perfectly implemented, limits the spectrum to 6 MHz bandwidth (5.38 MHz + 11.52%). A perfect filter, however, is not possible due to filter tap limitations and finite resolution. The filter is a complex filter to accommodate the vestigial sideband function. In addition, the low level VSB pilot is added to the signal at a constant RF level. The pilot aids carrier recovery and is used to remove phase ambiguity.

Careful attention must be given to signal peaks that arise due to the bandlimiting function of the Nyquist Filter. Some peaks exceed a 6 to 7 dB peak-to-average signal ratio. These peaks are generally clipped in the power amplifier stages to maximize PA efficiency. The effect of the clipping is not considered serious because of the rarity of large peaks. The key is to provide sufficient headroom in power amplifiers so that the spectral mask can be met with appropriate precorrection.

In-band equalization may be programmed into a DTV modulator using software to generate an equalization function unique to each transmitter at the time of manufacture. Equalization is applied by modifying the IF digital filter coefficients in the modulator which are then held in modulator memory as shown in Figure 6.2-34. This function can also be performed adaptively by digital signal processing (DSP) to assure optimum operation of the transmitter as operating conditions change. A companion demodulator is required to provide a sample of the output from which the DSP updates the coefficients as transmitter characteristics vary.

Upconversion & Final Amplification

The requirements for upconversion and final amplification for digital transmission are similar to those for analog signals. The IF frequencies selected will be different and there may be two stages of upconversion rather than one; however, the important performance characteristics of oscillator phase noise, frequency stability, and susceptibility to microphonics remain the same. As with analog visual and aural signals, it is important that the final amplifer have a linear transfer characteristic in amplitude and phase and flat, symmetric frequency response and minimum group delay variation across the modulation passband. For digital signals at least 6 MHz bandwidth is required.

For DTV transmitters operating on channels immediately adjacent to (above) and within 88 km of a transmitter on a NTSC channel, the pilot frequency must be maintained 5.082138 MHz above the visual carrier frequency of the NTSC transmitter within a tolerance of ±3 Hz. This requires that both carriers be locked to a common, highly stable source.

POWER AMPLIFIER DESIGN CONSIDERATIONS

Aside from near linear, transparent performance, it is essential that TV transmitters operate efficiently with minimum lost air time. Efficient power conversion, while important at VHF, is unusually important for UHF transmitters. This factor is expected to impact more stations as the DTV service comes on-line, pre-

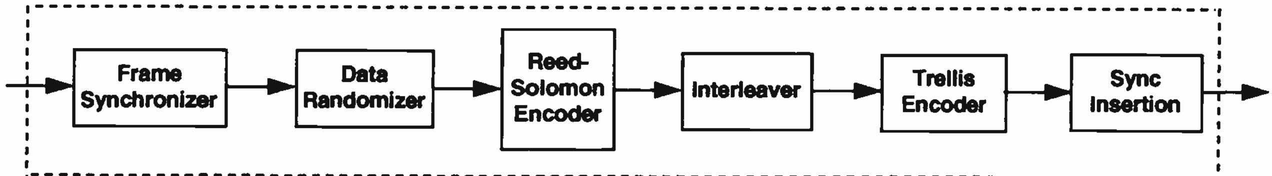


Figure 6.2-33. Transport to transmission conversion.

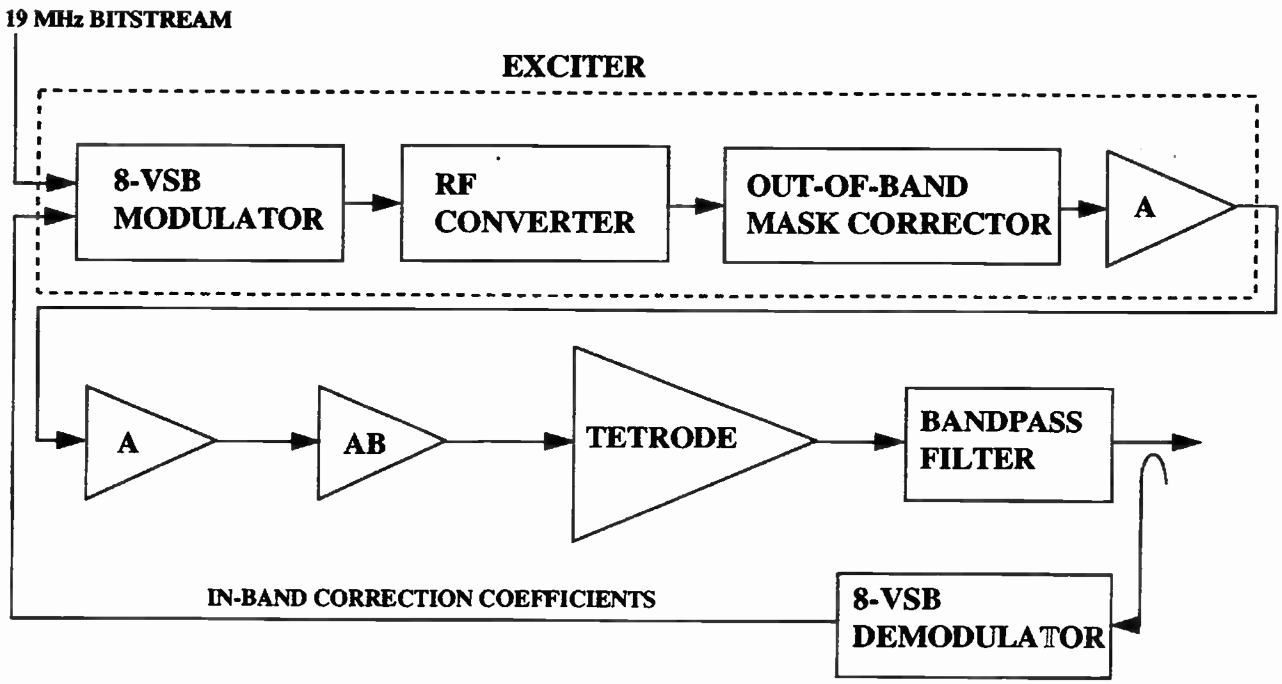


Figure 6.2-34. Block diagram of a tetrode DTV transmitter.

dominately in the UHF band. Many stations operate continuously unattended, making reliability a key requirement.

Reliability

There are many factors which can affect the reliability of a TV transmitter. Overall design philosophy, device technology, module design, control architecture, power supplies, cooling and cabinet design are critical areas which must be considered. Consider a transmitter design which uses circuits in series with no system redundancy; if one device fails, the entire transmitter fails. Figure 6.2-35 shows a system of three series devices with no redundancy. If each device (a, b, c) has a probability of survival (P) of 0.5 over some time interval the probability of the system surviving $P(s)$ is given by the formula:

$$\begin{aligned} P(s) &= P(a) \times P(b) \times P(c) \\ &= 0.5 \times 0.5 \times 0.5 \\ &= 0.125 \end{aligned}$$

If three identical devices are operated in parallel (Figure 6.2-36) and only one is required for adequate operation of the system, the probability of the system surviving is greatly enhanced. The overall system survival probability now becomes:

$$\begin{aligned} P(s) &= P(a) + P(b) + P(c) \\ &\quad - P(a)P(b) - P(a)P(c) - P(b)P(c) \\ &\quad + P(a)P(b)P(c) \\ &= 1.5 - 0.25 - 0.25 - 0.25 + 0.124 \\ &= 0.875 \end{aligned}$$

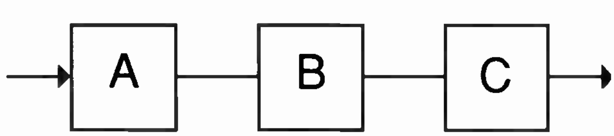


Figure 6.2-35. Circuits in series.

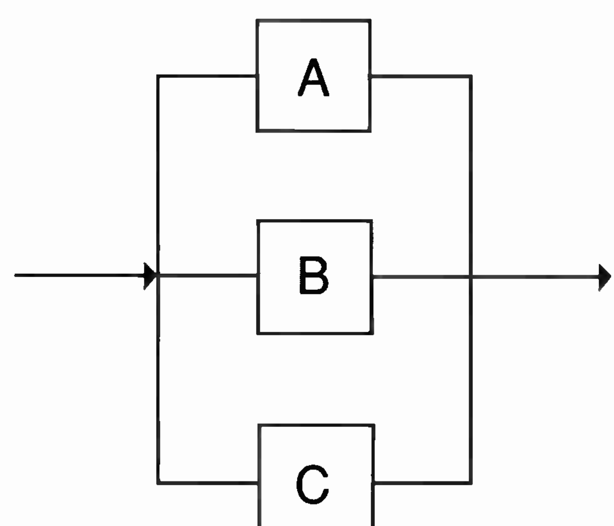


Figure 6.2-36. Circuits in parallel.

On-Air Availability

Related to reliability, but perhaps even more important, is on-air availability. On-air availability is the percentage of time the transmitter is available for service, defined by the following equation:

$$\text{On-Air Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR} + \text{MPMT}} \times 100\%$$

where:

MTBF = Mean Time Between Failures (hours)
MTTR = Mean Time To Repair (hours)
MPMT = Mean Preventative Maintenance Time (hours)

It is apparent there is little point in designing a transmitter that has a high MTBF figure if, due to poor design and mechanical packaging, it takes an inordinate length of time to make repairs, or the transmitter has to be shut down frequently for routine preventative maintenance.

Many stations have very short sign-off windows or operate 24 hours a day. This often results in a less-than-optimum maintenance schedule which can lead to premature failure or out-of-tolerance operation. One way to reduce the amount of off-air maintenance time is by making provisions for on-air maintenance or to have redundant transmitters. This significantly reduces the MPMT.

Several design factors should be considered for optimum on-air availability:

- High reliability for the fundamental circuits
- Provision for fast, easy access to all subassemblies
- Maximum use of like parts and subassemblies. Because fewer items are needed, this allows a TV station to maintain a full inventory of spares. If spares are on hand, it follows that the repair time is much shorter
- Repair of transmitter at subassembly level. A subassembly which has been removed can then be repaired by station personnel or returned to the manufacturer for exchange.

Efficiency

When the fundamental concept of efficiency is applied to TV transmitters several factors must be considered. For purposes of determining total power consumed, the ac to RF conversion efficiency is the parameter of interest. For systems with unity power factor, determining the ac input power is relatively straightforward. For power factors less than unity, the relative phase of the fundemental voltage and current must be determined. In addition, in systems generating significant line harmonics, the relative level of these harmonics must be known. Power factor is expressed either as displacement power factor or total or true power factor.

Displacement power factor is is the cosine of the phase between the voltage and current at the fundamental frequency. It is equal to the total power factor only for an undistorted sine wave.

Total power factor is the ratio of total power to the apparent power, given by:

$$\text{PF} = \text{ac power input (watts)/V (volts) I (amps)}$$

Determination of input power is somewhat simplified by considering only the dc to RF conversion process. In this case it is necessary to determine the voltage and current provided by the power supply(s) for the final amplifiying devices. While this is a useful tool for comparison purposes, it has the disadvantage of ignoring the power consumed elsewhere in the transmitter such as driver stages, cooling systems, filament and magnet power and control. If these powers are to be included in the efficiency calculation, they must be determined separately.

For NTSC transmitters, determination of output power is equally complex. Transmitters are commonly rated in terms of peak sync visual power. Exclusive use of this number neglects the aural output. In addition some amplifier technolgies may exhibit efficiencies greater than 100% using visual peak power. This has given rise to the use of figure of merit defined as:

$$\frac{\text{Visual peak sync output power}}{\text{dc input power at 50\% APL}}.$$

This definition is valid for transmitters using separate amplification. For common amplification, the aural output and input powers must be added to the numerator and denominator, respectively. Typical figures of merit for common tube amplifiers operating in separate amplification are shown in Table 6.2-1.

Consider a typical klystron amplifier in pulsed operation. Direct current input power is calculated as follows:

$$\begin{aligned}\text{DC input power} &= \text{Beam voltage}\\ &\quad \times [(\text{Beam } I_{sync} \times \text{Sync duty cycle})\\ &\quad + (\text{Beam } I_{video} \times \text{Video duty cycle})]\end{aligned}$$

For 60 kW output, dc input power = 24[(5.5 × .08) + (3.7 × .92) = 92.2 kW

$$\text{Figure of merit} = 60/92.2 = 0.65$$

For digital transmitters, the situation is somewhat simpler. The output power most readily measured is the average level which is constant. Since the audio

Table 6.2-1
Typical figures of merit for common tube amplifiers operating in separate amplification.

AMPLIFIER DEVICE	FIGURE OF MERIT
TETRODE	.9–1.0
INTEGRAL CAVITY KLYSTRON	.65–.75
EXTERNAL CAVITY KLYSTRON	.65–.75
KLYSTRODE OR IOT	1.1–1.3
DEPRESSED COLLECTOR KLYSTRON	1.2–1.3

is encoded with the video, there is no separate aural amplifier. Thus the dc to RF efficency is readily calculated as a ratio of the average RF output to the dc input.

SOLID-STATE TRANSMITTERS

Technological advances in bipolar and field effect transistors (FET) have made the development of solid-state high power, linear amplifier modules for TV applications both practical and cost effective. By combining RF modules, it is practical to create transmitters at any power range up to 75 kW. Solid-state transmitters maintain their performance over extended periods of time due primarily to the fact that they have no tuning controls nor filament emission degradation with time. No warm up time is required—solid-state transmitters produce full rated power within seconds of activation.

Solid State Devices

Both bipolar and FET technology exist today as suitable RF amplification devices. Power amps are operated in class AB for the best trade-off of efficiency and linearity. Driver stages usually contain Class A amplifiers.

Although both device types have merit, FETs have some advantages over bipolar devices. FETs have a higher amplification factor than bipolar transistors, helping to reduce the number of driver stages. Higher supply voltages help to reduce the current capacity of the power supply. Simpler bias circuitry minimizes parts count.

RF Amplifiers

Combining several RF power modules to achieve the desired transmitter output power increases the parallel redundancy and the on-air availability. Output power of 1–2 kW per module has been adopted by nearly all manufacturers based on overall cost, practical weight, and size limitations.

Self protection of each PA module against various fault conditions is good engineering practice. By using self protecting modules, the cabinet control logic and overall transmitter control logic can be kept simple, thus improving overall reliability. Self diagnostics for the module aid in minimizing time to repair. Protection from over-voltage, overdrive, VSWR, overtemperature and ensuring proper load sharing among devices is essential to maintaining amplifiers for long life. It is desirable that one subassembly failure not cause another subassembly to fail.

Modular amplifiers which can be removed while hot improve overall on-air availability. If an amplifier module fails, the transmitter continues to function indefinitely without disrupting transmitter operation. If a spare PA amplifier is on hand, it can be used while the failed unit is repaired.

Temperature compensated regulated supplies for the amplifiers are important. Otherwise, power output varies as temperature or supply voltage changes.

Combiners and Dividers

There are several methods used for combining RF power for the solid-state amplifier modules. An effective method is the use of in-phase N-way combiners. Three common examples are described.

Microstrip Wilkinson Combiner

Figure 6.2-37 shows an example of a Microstrip Wilkinson Combiner. Microstrip is used to carry the RF power. When all amplifiers are operating, equal voltages are presented to each node of the load resistor so that no power is dissipated. When an amplifier failure occurs, the power is distributed between the load and the output. The impedance of the transmission lines and the length of the lines is selected to achieve the desired impedance transformation. Balanced reject loads are used to absorb RF power in case of amplifier failure and to provide isolation between amplifiers.

The combiner is housed in a shielded casing so that outside fields cannot alter the balanced configuration. The reject loads are cooled by conducting heat through the flange to the heatsink where the heat is exchanged to the moving air stream. This type of combiner is used generally for lower numbers of amplifiers (two to six).

Ring Combiner

Figure 6.2-38 shows an example of a ring combiner. The higher power handling capability of the coax lines used in this combiner allows a range of amplifiers to be combined (two to 20). It also provides isolation between amplifiers using reject loads which are not in the direct RF path to the output. The operation of the multiport combiner is easily understood by considering a two-way version (see Figure 6.2-39). Each transmission line is a quarter wavelength. When equal voltages are applied to both input ports (both amplifiers operating), the combined signal arrives at the output. This is true because the distance from each input port to the output is electrically equal whether the signal follows the shorter or longer path, the signal from one amplifier arrives at the load port out of phase with that

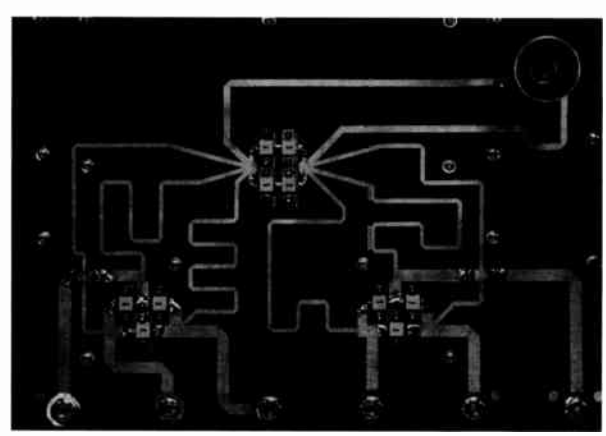


Figure 6.2-37. Microstrip Wilkinson combiner (Photo courtesy of Larcan).

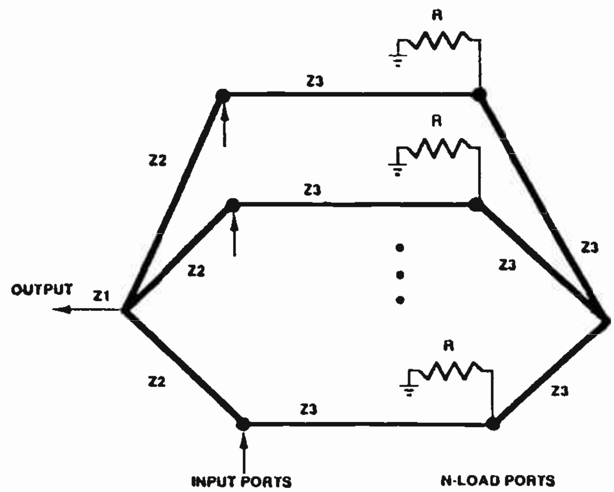


Figure 6.2-38. N-way ring combiner.

from the other amplifiers so no power is dissipated and the signals from amplifier 1 arriving at input port 2 via the short and long paths are out of phase and vice versa. *Under normal conditions, all power appears at the output, none is absorbed in the loads, and there is complete isolation between amplifiers.*

Power is absorbed in the load resistors only when an amplifier is not operating. Assume that only amplifier number 1 is operating. The signal path is electrically equal not only for the long and short paths to the output, but also, for the right and left paths to either load. The power from the operating amplifier is split equally between the output and the isolation loads. Due to the isolation inherent in the network, the input ports remain matched even when one or more amplifiers are removed.

Since the transmission line that connects the amplifier port to the output port is a quarter wavelength long and 75 Ω transmission line is used, the 50 Ω amplifier impedance is transformed to: $(75/50) \times 75 = 112.5$ Ω. For combining N amplifiers, the impedance at the output combiner junction is 112.5/N. This impedance is then matched to 50 Ω.

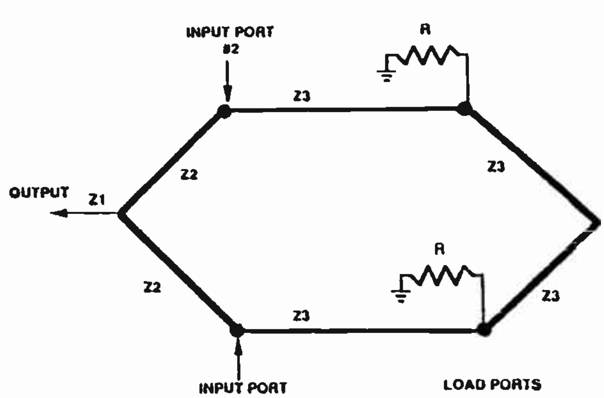


Figure 6.2-39. Two way ring combiner.

The reject loads are mounted to a grounded structure. One method of cooling the loads is to use a heat pipe for the mounting structure. In case of amplifier removal, the reject load temperature rises, the fluid in the lower section of the heat pipe heats until it vaporizes. The vapor rises to the finned area where the heat is exchanged to the moving air stream. The vapor condenses as it releases its heat and returns to the bottom of the pipe to repeat the cycle.

Starpoint Combiner

Starpoint combiners are simple low loss devices and operate in the following manner. Consider parallel operation of four sources of equal phase and amplitude, each with a source impedance of 50 Ω and no interaction. The resulting source impedance is 12.5 Ω. All that is needed to complete the combining process is to transform this impedance to 50 Ω. Obviously, this concept can be extended to any value of "N." The "Q" of the impedance transformation determines the combiner bandwidth. For optimum bandwidth it may be necessary to combine and transform in several corporate steps. In reality, the combined amplifiers have slightly differing phases and amplitudes. As in any combiner design, this represents a combining loss that must be accounted for. A flat air spaced stripline structure may be used to provide a combiner with extremely low resistive loss.

This simple combining technique works best only if all inputs are present. If a failure occurs the isolation may be insufficient to prevent interaction between modules, especially for small values of "N." Using circulators at the output of each module resolves this, producing a combiner with excellent isolation and bandwidth. Circulators not only provide port-to-port isolation but also protect against high VSWR due to icing and other causes. The combiner band split should be consistent with module and circulator bandwidth.

Combining Multiple Amplifier Cabinets

When combining RF power amplifiers, they must be matched in phase and gain for maximum power to the antenna and minimum power to the reject load. Electronic phase shifters and attenuators should have the capability for remembering their settings in case of ac power failure.

Cooling System

Proper cooling of the solid-state modules is important for high MTBF. The MTBF of a transistor essentially doubles for every 10°C drop in the junction temperature. Distributed cooling systems employing more than one fan offer good redundancy. Current motor/fan technology has matured to the point where a few larger direct drive fans are as reliable as many smaller fans. Since many RF power amplifiers may be employed, a large volume of air is needed to adequately cool the heatsinks. Low pressure fans or blowers may be used if heatsink fin density is not high. This aids in reducing audible noise. The heat is distributed over a large volume of air and the temperature rise is relatively low.

Automatic Gain Control (AGC)

AGC is used to maintain constant power output from the transmitter. Ambient temperature changes will cause gain changes in a solid-state amplifier. RF drive power must be changed to maintain constant power output. A detected RF sample of the PA output may be fed to an input of a comparator. The exciter output sample or a voltage proportional to the exciter output is applied to the other input of a comparator. The dc output is then integrated and fed to an attenuator which varies the RF drive level at a low power level. Alternatively, the RF sample may be taken from an intermediate stage. In this case, other means must be used to temperature compensate the output stage or some power reduction with temperature must be tolerated.

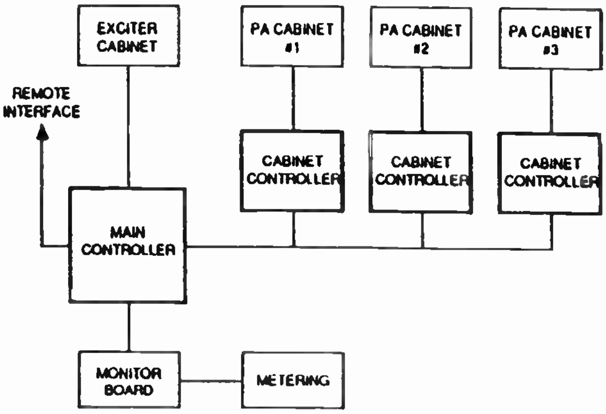


Figure 6.2-40. Distributed control and monitoring.

Power Supplies

Power supply design is critical to the reliability of a solid-state transmitter. Since FET and bipolar devices are low voltage devices, the power supplies which serve them must provide low voltage and high current. High reliability connections must be guaranteed in the dc distribution. Since available power output from a transistor varies as

$$Po = \frac{(V)^2}{2R_L}$$

it is desirable that the supply remain very tightly controlled over incoming ac line variation. Since the amplifier current demand varies with picture level (for NTSC), the power supply output voltage must remain stable from low load (white picture) to high load (peak sync output). Efficiency of the power supply is important since the lost power results in heat as well as unnecessarily high utility costs. Any voltage or current transients should be significantly suppressed before reaching the amplifier transistor device. Transmitters should successfully pass the applicable portions of the ANSI/IEEE C62.41 Transient Testing Standard (also referred to as the IEEE-587 Standard).

Alternating Current Distribution

A reliable method of ac distribution provides power to modular RF amplifier cabinets through a parallel system. Each cabinet is protected by a separate ac breaker external to the transmitter. This concept allows a cabinet to be safely serviced while the remaining cabinets are operational. Phase monitors guard against low voltage, loss of phase or reversal of phase.

Control Systems

If individual amplifier modules and power supplies are self protecting, control and monitoring functions can be simple and straightforward. One approach for the control system is to use a single controller to control and monitor all the functions of the transmitter. Another approach is to distribute the control system throughout the transmitter. The distributed control system can be designed so that the failure of any individual controller does not affect the operation of the others.

Both distributed control and centralized control systems are shown in Figures 6.2-40 and 41. After ac power failure, the controller should have backup memory to restore the transmitter to the same operating condition as before.

A system of indicators is essential to quick fault diagnosis. Typical status conditions which may be displayed are: exciter fault, VSWR fault, VSWR foldback, power supply fault, controller fault, air loss, door open, fail-safe interlock, phase loss, module fault, visual drive fault, aural drive fault and external interlock(s).

VSWR foldback reduces power during high VSWR operation, such as antenna icing, and restores RF power to normal when difficulties are removed. Other options used to enhance the on-air capability may include dual exciters, 20% aural power and redundant drive chains. A block diagram of a solid-state transmitter with VSWR foldback is shown in Figure 6.2-42.

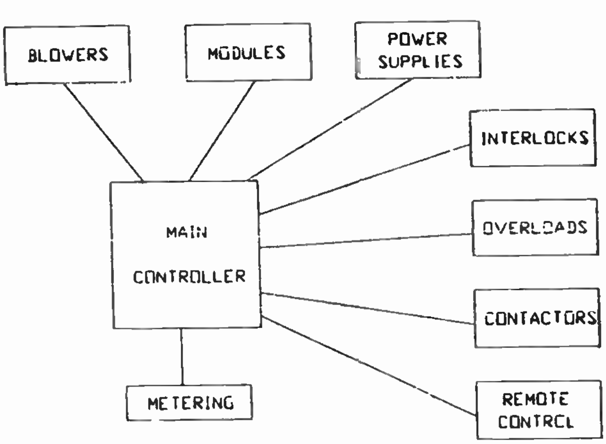


Figure 6.2-41. Centralized control system.

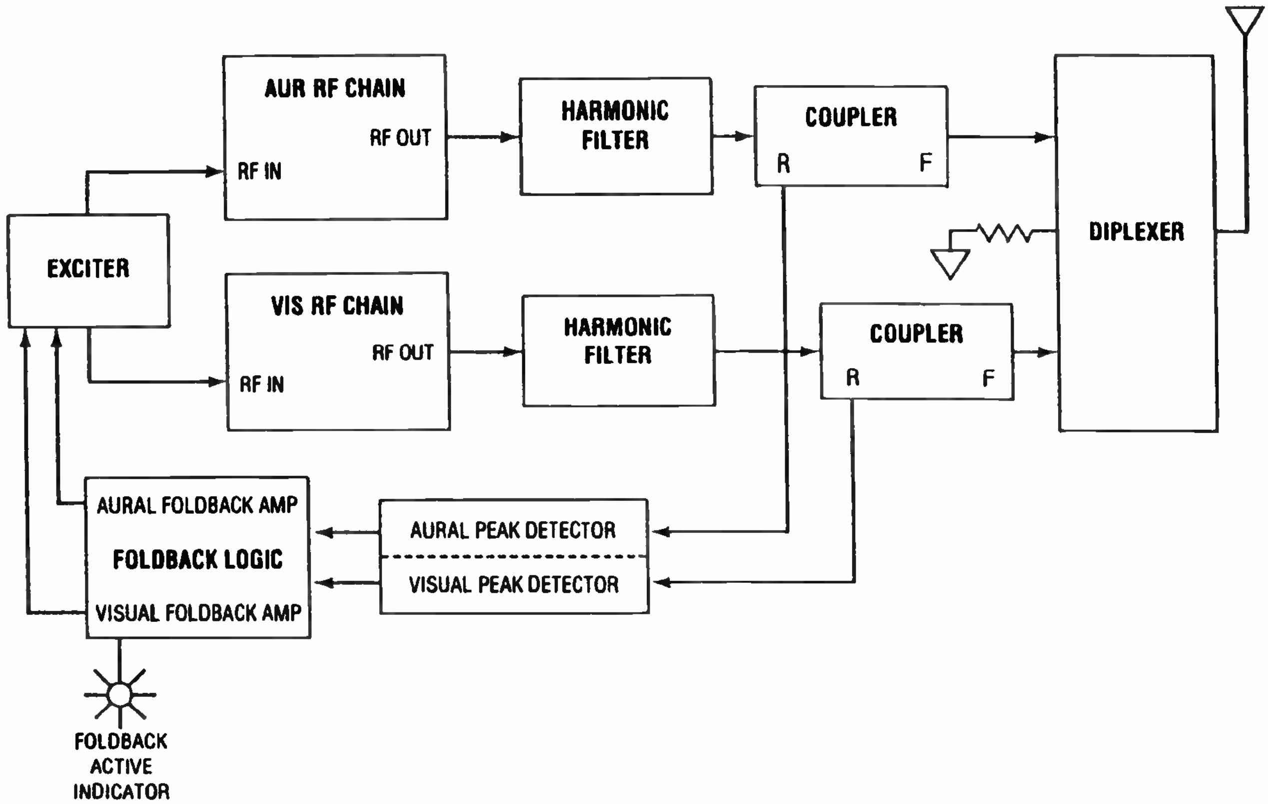


Figure 6.2-42. VSWR foldback block diagram.

UHF TUBE TRANSMITTERS

Due to higher transmitter power, power consumption and efficency are of utmost importance for UHF transmitters. A variety of tube technologies are available to address this requirement. Some are most suited for lower power transmitter designs, others are more appropriate for the highest power requirements. These technologies include tetrodes, klystrons, multiple depressed collector klystrons (MSDC) and inductive output tubes (IOT).

Tetrodes and Diacrodes

Tetrode refers to a generic category of four element tubes suitable for the linear amplification of RF signals to the power levels for required for broadcasting. This discussion is limited to water cooled UHF TV tubes made available in recent years. These tetrodes are of cylindrical construction. The inner diameter of the anode of a typical 30 kW peak sync tetrode is barely 2 inches in diameter. The anode is domed at the upper end to make it watertight. The water jacket is a watertight container around the anode. Element spacing is quite close, especially the spacing between the control grid and the cathode which is barely .003 in. Gain is an inverse function of the control grid to cathode spacing while a direct function of control grid to anode spacing. Figure 6.2-43 is a horizontal cross sectional sketch of a typical power tetrode.

Tetrodes are biased for Class AB operation and are, therefore, more efficient than the Class A klystron. Tetrodes operating at UHF at higher power are made possible due to advances in ceramics and ceramic-to-metal bonding, pyrolytic grid structures and the method of anode cooling. They exhibit excellent envelope and RF linearity requiring only the degree of correction normally seen in VHF transmitters. The tradeoff for performance in these areas is lower power gain than most other amplifiers. Gain for a tetrode is about 15 dB.

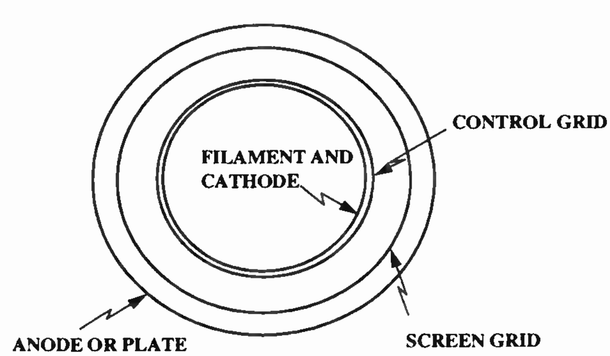


Figure 6.2-43. A horizontal cross section of a typical power tetrode showing the relative placement of the four elements.

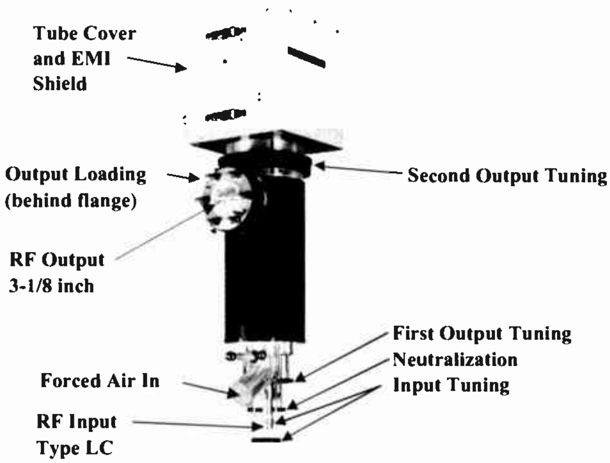


Figure 6.2-44. A representative cavity for the tetrode is shown. RF drive connects to the bottom while the output is taken from the side through the 3-1/8th in. EIA flange.

Modern UHF power tetrodes for linear television applications are capable of combined visual and aural amplification. Ten percent aural power is the norm and tubes are rated according to peak sync power with the aural carrier. For example, a typical UHF tetrode may be rated at 30 kW peak sync. Ten percent aural power is assumed giving a PEP rating of 52 kW. It is capable of 43 kW peak sync if the aural carrier is amplified separately.

Connections to these tubes are made at the base by means of concentric ring conductors, staggered in height and mating to a socket which sits above the tuning sections of the input and output networks. This allows nearly all the RF current to flow in the lower two thirds of the tube's vertical structure, the upper end being open circuited. Since the tube sits above the cavity sections, tuning is not disturbed when the tube is removed or inserted. Because inner electrode capacitances vary from tube to tube, some touch-up retuning is required. Tuning is not done at full power but at milliwatt levels.

Typical tetrodes are operated in grounded grid. Neutralization is required, but it is a one time physical setting and does not change from tube to tube. The tetrode is cathode driven with a low "Q" input tuning section. Output power is taken from the anode through a double tuned cavity section with loading control. The grids are grounded for RF. The tubes are cooled with distilled water. Cooling water must be non-conductive since it comes in contact with anode voltage. A representative cavity is shown in Figure 6.2-44.

The tetrode node voltage is much lower than that of the klystron. No special standoff requirements are necessary and the negative supply line of the power supply is grounded. The cathode is at dc ground potential.

Contributing to the service life of modern tetrodes is the water cooling of anode and screen grid. The method used is a process in which the water coming in contact with the anode surface is allowed to vaporize to the *nucleate boiling* state within tiny pits on the anode surface as illustrated in Figure 6.2-45. As soon as the vapor enters the main water stream a few thousandths of an inch above the surface, it condenses, releasing its heat content to the main stream. Nucleate boiling is enhanced by deliberately roughening the anode surface providing millions of boiling sites uniformly spread over the surface. If the anode surface were visible and not hidden under a watertight cap, nucleate boiling would be recognized by the formation of nearly microscopic bubbles, which quickly condense and disappear. Film boiling is maintained in between nucleate boiling sites.

The screen grid is also water cooled, but no boiling takes place.

The important electrical characteristics of a 30 kW peak sync tetrode are shown in Table 6.6-2. The basic structure of these tubes is silver-plated copper and ceramic with pyrolytic graphite grids and a thoriated tungsten directly heated cathode. Filament currents are high to minimize cathode current modulation of the cathode temperature.

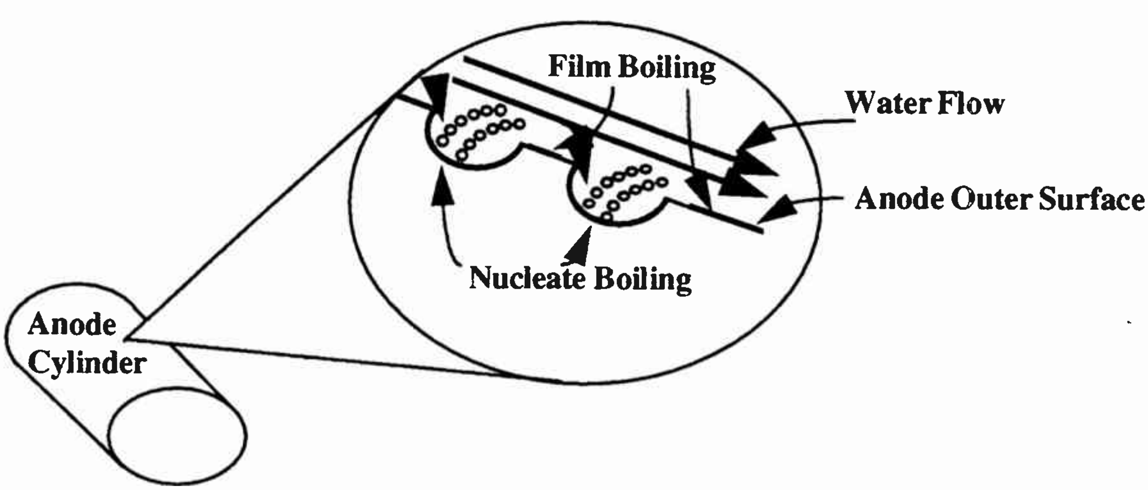


Figure 6.2-45. Cooling of the anode surface.

Table 6.2-2
Typical power tetrode electrical characteristics.

$P_{out\ PEP}$	52 kW
$P_{out\ peak\ sync}$	30 kW
10% Aural?	Yes
One dB Bandwidth	12.8 MHz
Circuit Configuration	Grounded Grid
Integral to the Cavity?	No
Power Gain	15 dB
IMD	−48 dB no correction
Neutralization	Yes
Sync Compression	5 IRE no correction
Plate (Anode) Voltage	8500 V
Anode Current Idle	2.5 A
Anode Current Black	5.6 A
Screen Grid Voltage	650 V
Screen Grid Current Black	40 mA
Control Grid Voltage	−120 V
Control Grid Current Black	20 mA
Filament Voltage	5.1 V ac or dc
Filament Current	180 A

Drive power is in series with the input/output circuits which are in series. Since the cathode is the driven element, its current is nearly identical to the anode current. Thus power gain is realized from the voltage gain.

The *Diacrode®* is also basically a four element tetrode. The same construction is used even though the shape is nothing like that of the tetrodes. Figure 6.2-46 shows a three dimensional cutaway drawing. It is made of the same materials as the tetrodes and fits into the same socket in the same cavity modified only slightly to accommodate the extra height for an external element. The operating voltages are the same except for the filament, but it delivers twice the power. (Diacrode® is a registered trademark of Thompson Tubes Electroniques.)

The flat shape allows a geometrical modification of the anode structure so that the top of the tube may be fitted with an external circuit element electrically connected between the anode (8500 VDC) and screen grid (600 VDC). The RF current and voltage distribution is shown in the drawing at the left of Figure 6.2-47.

The external circuit element is a high "Q" cylindrical coaxial transmission line electrically RF connected between the anode and screen grid. It has an insulating layer of .005 in. kapton between the cylinders used as a dc block.

Along the length of the cylinder is a circular sliding RF short circuit between the two cylinders. The purpose of the upper high "Q" transmission line section, tuned to be a shorted quarter wavelength is to reflect an open circuit to the vertical center of the anode. This causes the vertical open circuit or current minimum within the tube to be at the vertical halfway point. The result is a voltage maximum at the same point. Both vertical halves of the tube are used to make power. Since there are two current maximums, one at the usual place at the base and one at the sliding short in the upper transmission line, anode current is doubled, but operating voltages remain the same as the tetrode. The setting of the upper cavity section is quite broad with only three settings to cover the UHF band from channel 14 to channel 69.

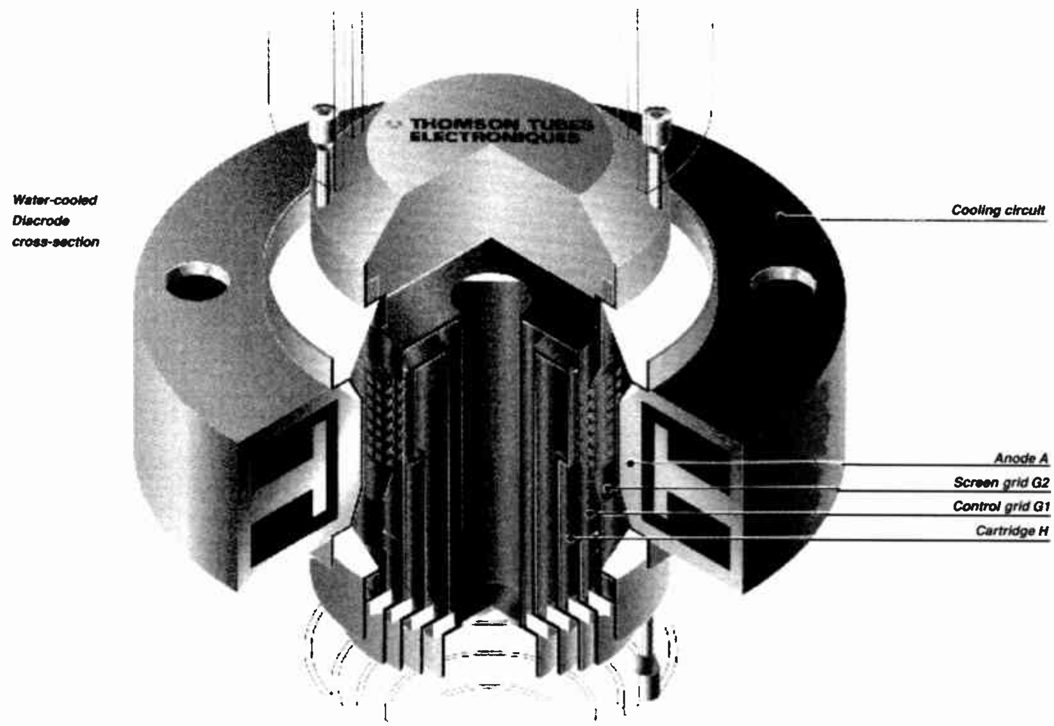


Figure 6.2-46. A three dimensional cutaway of a Diacrode® showing all four elements of the tube and anode water path.

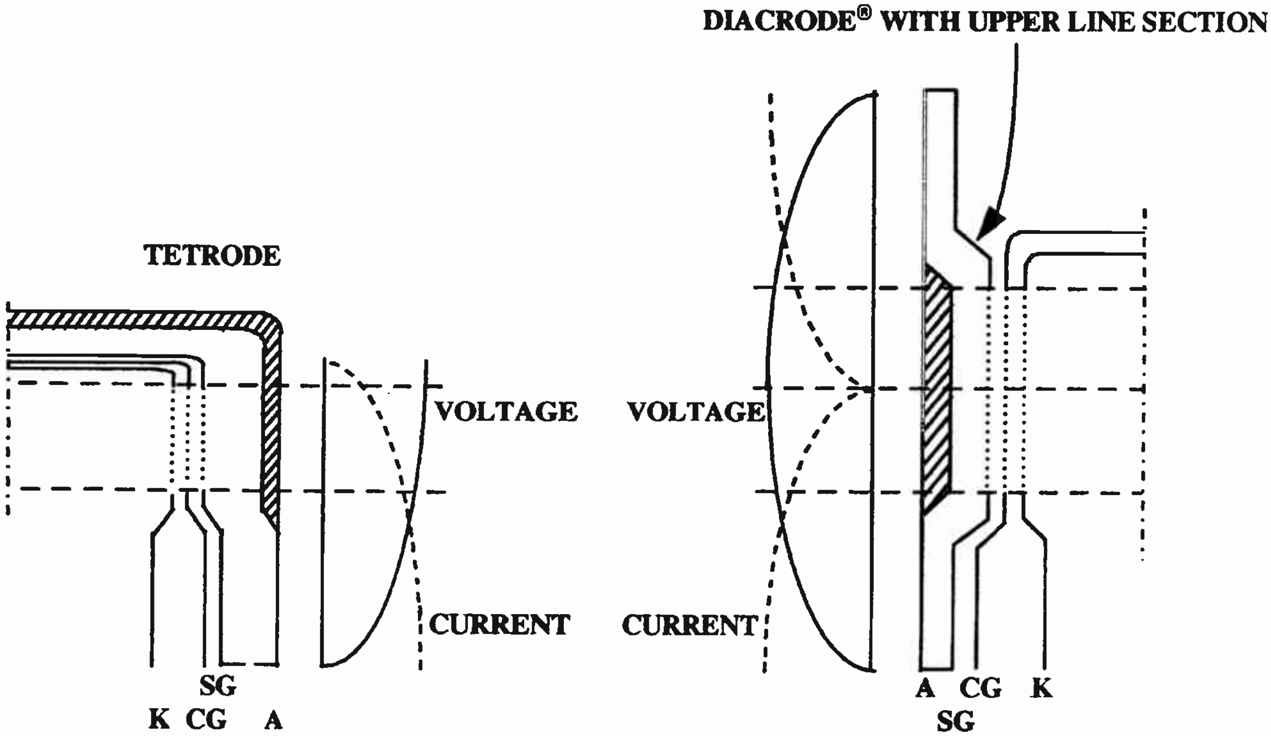


Figure 6.2-47. The upper shorted quarter-wave line section above the Diacrode® causes two current maximums instead of the single one in the tetrode.

The Tetrode Transmitter

A block diagram of a NTSC tetrode transmitter using a 30 kW tetrode is shown in Figure 6.2-48. It is an IF modulated combined amplification transmitter requiring almost 1 kW of drive power. Drive is supplied by two solid-state Class AB amplifier drawers having gain of about 23 dB. Correction circuits include low frequency linearity, differential gain, differential phase and sync compression, all performed at video and ICPM, stereo pilot protection, in-band intermodulation distortion and cross modulation, all performed at IF.

For good linearity the control grid and screen grid power supplies must be well regulated. These supplies are capable of modulating the amplified signal if allowed to shift with modulated drive level. It is not necessary to regulate the anode supply since the tetrode is not saturated and is relatively insensitive to anode voltage variations. Both oil filled and dry supplies may be used for the anode voltage source. Oil has a cooling advantage over convection or forced air cooling.

The filament supply must also be well regulated and should be capable of two different voltages: full voltage according to the value in Table 6.2-2 and about 20% of full voltage. The latter is known as *black heat* voltage and is used when the tube is in off air, standby mode. No cooling of any kind is needed during black heat.

Black heat is used to reduce the time to on-air, since the filament is already warm, and to reduce the thermal stress of going to full filament power. The filament voltage should be gradually increased from the black heat value to full value over a period of 60 seconds. Likewise, it should be gradually decreased at the end of the broadcast day for longest life.

It is imperative that the control circuitry be designed so that faults, over-voltage and over-current situations quickly disconnect the power supplies from the tube. It is also imperative that the cooling system be maintained in proper operating condition. Blocked air filters or reduced water flow due to blockages will shorten tube life even though the control system judges conditions are adequate to operate. The control system should be designed so that the transmitter will remain on air as long as minimum cooling requirements are met. The transmitter should be designed to have warning indicators posted on the front panel, available by remote monitoring. When a trip occurs, the high voltage power supply should be removed within 15 msec and dc output of the screen grid power supply immediately shorted to ground while being disconnected from the ac mains.

Klystrons

The *klystron* uses velocity modulation in the amplifying device. The electron beam emitted from the cathode is accelerated to high velocity by the electric field between the cathode and anode and is directed into the RF interaction region, as shown in Figure 6.2-49. An external magnetic field is employed to focus

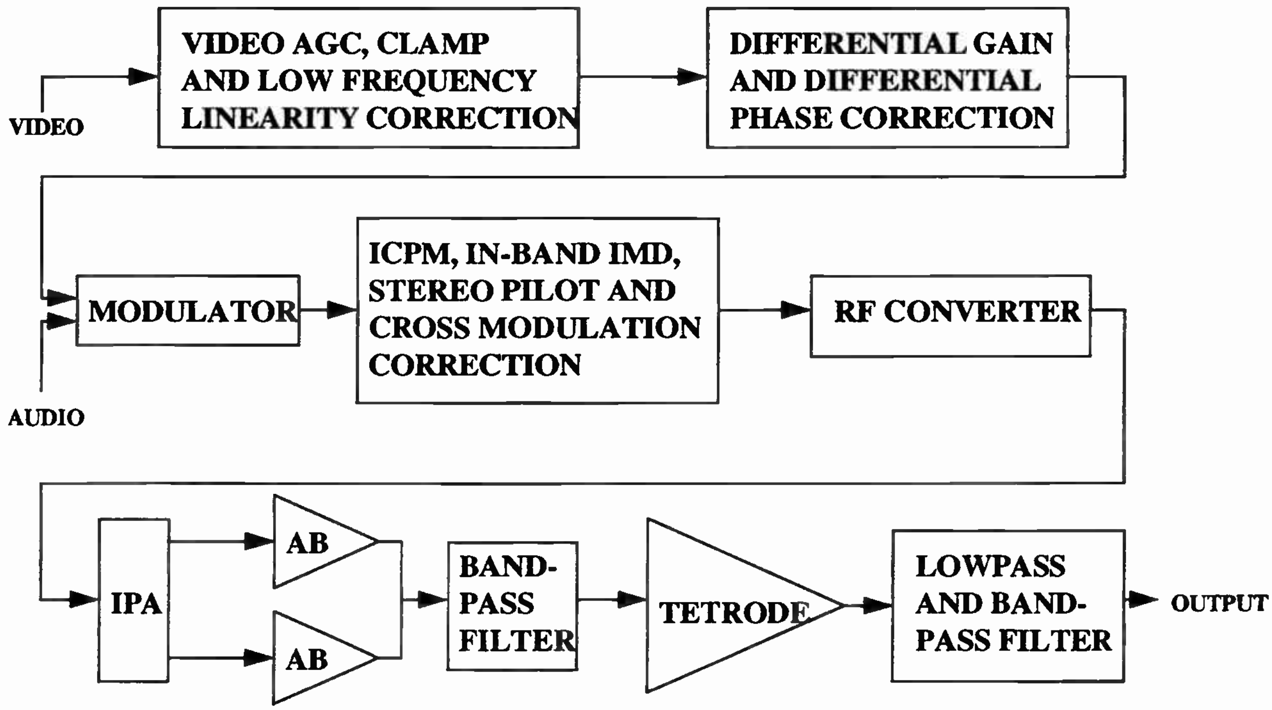


Figure 6.2-48. Block diagram of a 30 kW common amplification tetrode transmitter.

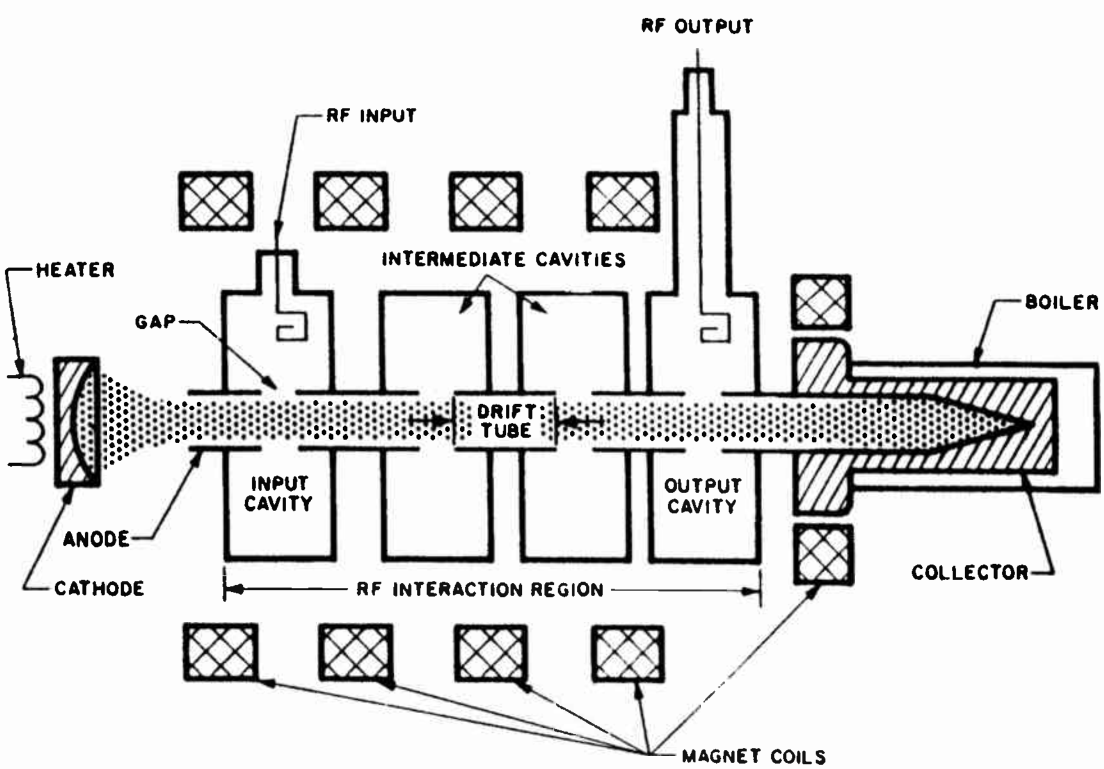


Figure 6.2-49. Principal elements of a klystron.

the beam as it passes through the tube. The electron beam impinges on the collector, which dissipates the beam energy and returns the electron current to the beam power supply.

The RF interaction region, where the amplification occurs, contains resonant cavities and field-free drift space. The first resonant cavity (the input cavity) is excited and a RF voltage is developed across the gap. Since electrons approach the input cavity gap with equal velocities and emerge with different velocities, the electron beam is said to be *velocity modulated.* As the electrons travel down the drift tube, bunching develops and the density of electrons passing a given point varies with time.

The RF energy produced by this interaction is extracted from the beam and fed into a coaxial or waveguide transmission line by means of a coupling loop in the output cavity. The dc beam input power not converted to RF energy is dissipated in the collector. The cavities can be mounted external to the klystron or can be included in the vacuum envelope. The resonant frequency of each of the cavities is adjusted by changing the volume of the cavity in external cavity klystrons, or the capacitance of the drift tube gaps is changed in integral cavity klystrons.

The RF energy is fed through a coaxial line with its center conductor inserted into the cavity. The end of the center conductor is formed into a loop. This is a simple one-turn transformer which couples energy into the cavity as shown in Figure 6.2-50. Intermediate cavities may have their loops coupled into RF loads to vary the "Q" of the cavities thus to change the overall bandpass characteristics.

All cathodes have an optimum operating temperature. The operating temperature of the cathode must be high enough to prevent variations in heater power from affecting the electron emission current (beam current). However, the temperature of the emitting surface must not be higher than necessary, since excessive temperature can reduce cathode life.

There are two electrodes which may control the beam current: the modulating anode and a lower voltage (0 to 1,400 V) electrode typically used for pulsing the beam current. If the low voltage electrode is connected to the cathode, the modulating anode voltage controls beam current. Perveance is a function of the geometry of the cathode-anode structure. This can be calculated using the following equation:

$$I_b = KV_b^{3/2}$$

where:

K = Perveance
I_b = Beam current in amperes
V_b = Beam voltage

Figure 6.2-51 shows the relationship between beam current and voltage described in the previous equation. In example A, if a modulating anode of 4,000 V with respect to the cathode beam voltage produces a beam current of 0.6 A, the intersection point lies at a perveance expressed as 2.4×10^{-6} or 2.4 micropervs. Operating condition B illustrates a practical television transmitter situation in which a common beam supply of 18 kV is used to power both the visual and aural klystron. The visual tube operates at a beam current of approximately 5.0 A if the modulating anode is connected (through an isolating resistance) to the body of the tube, and the perveance is 2.1 micropervs. Since the aural output power required is much less, the dc input power is less than that required to operate the

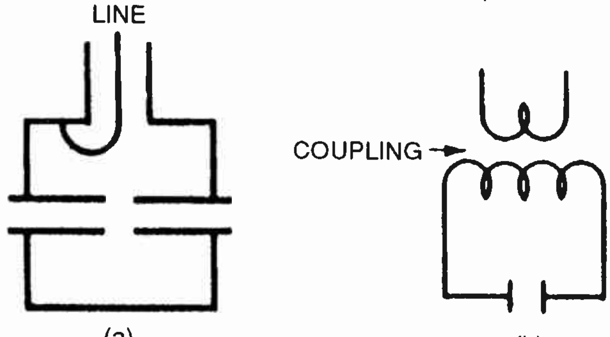


Figure 6.2-50. Loop coupling and equivalent circuit.

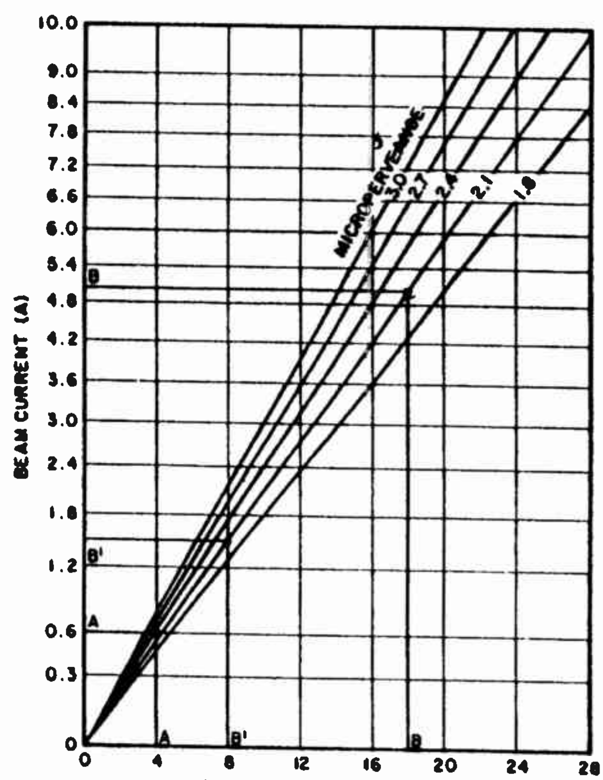


Figure 6.2-51. Beam current variation with modulating anode voltage.

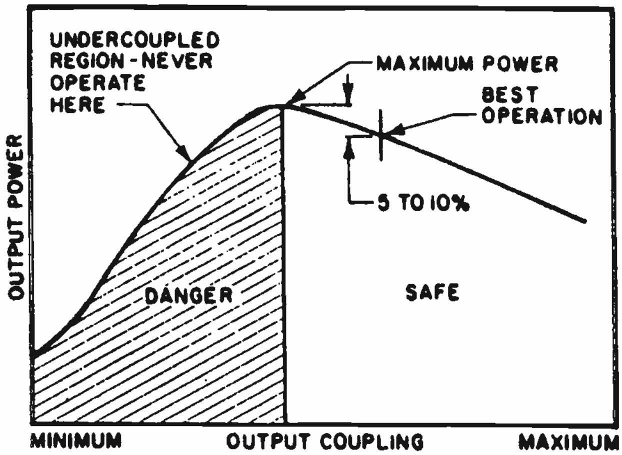


Figure 6.2-52. Adjustment of output coupling control.

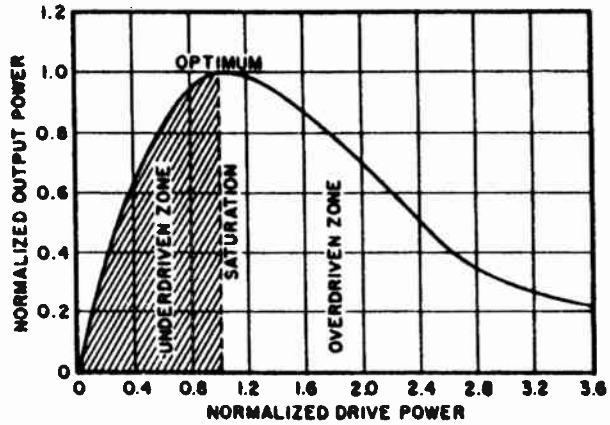


Figure 6.2-53. RF output power as a function of RF drive power.

visual tube. Points B′ indicate that if the modulating anode is supplied with only 8 kV (through a voltage divider), the intersection with the 2.1 microperv line yields a beam current of only 1.5 A, thus accomplishing the necessary reduction of input power for aural service.

Electromagnets are placed around the klystron to develop a magnetic field along the axis of the electron beam. This controls the size of the electron beam and keeps it aligned with the drift tubes. If the magnetic field is interrupted or insufficiently controlled, the electron beam will land on surfaces other than the collector and may destroy the tube.

The output cavity is generally tuned to the carrier frequency. It is essential to operate with the coupling loop adjusted so the output cavity is slightly over coupled. Figure 6.2-52 shows the relationship of output power to proper coupling loop adjustment. If the coupling loop is adjusted so that the cavity is under-coupled, arcing and ceramic fracture resulting in klystron failure may occur.

Figure 6.2-53 shows output power as a function of drive power. As expected, when drive power level is low, output power is low. As the level of drive power increases, output power increases until an optimum point is reached. Beyond this point, further increases in drive power result in less output power. In the zone labeled *Underdriven,* output power increases when the input power is increased. The point labeled *Optimum* represents the maximum output power obtainable. The tube is said to be saturated at this point, since any further increase in drive only decreases the output power. The zone formed at the right side of saturation is labeled *Overdriven.* To obtain maximum output power, sufficient drive power must be applied to the tube to reach the point of saturation. Operating at drive level beyond saturation point only overdrives the tube, decreasing output power, and increasing the amount of beam interception at the drift tubes (body current). Klystrons tuned for TV service are operated within the underdriven zone.

Figure 6.2-54 shows how output power changes with various levels of drive power applied under different tuning conditions. Point A represents saturation for a synchronously tuned tube. Points B and C show saturation reached by tuning the penultimate (closest to the output cavity) cavity to higher frequencies. Increasing the penultimate cavity frequency beyond Point D no longer increases output power. Instead, it reduces the output power as shown at Point E.

The input coupling is adjusted for the best tradeoff of minimum reflected power and best overall bandpass.

Operating a klystron at saturation improves efficiency but requires more linearity and phase compensation. By using a pulser to switch to a higher beam current during sync and back to a lesser current during video, the average beam current is significantly reduced. The practical limit of reduction of the video beam current is the point at which tip of burst and back porch distortion is not correctable.

With beam current pulsing, the effective "Q" of the electron beam and cavity combination is altered. This results in passband tilt from visual carrier down to the upper edge of the passband. This requires readjustment

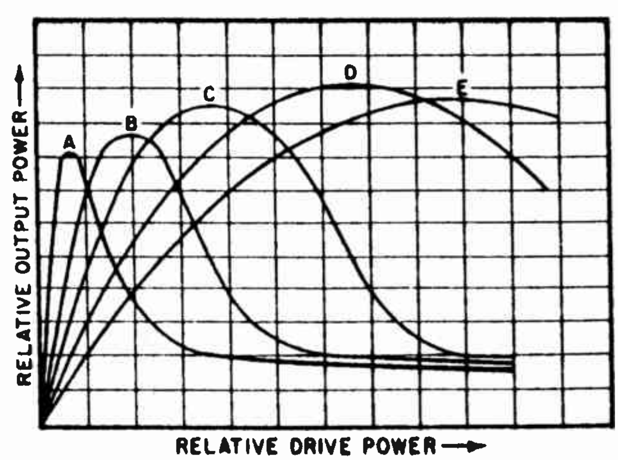


Figure 6.2-54. Output power variation with drive power under different tuning conditions.

of the cavities to obtain a flat response. Tilt of as much as 7 dB has been observed.

In static (nonpulsed) operation, the tube is supplied with enough beam current to saturate at 100% power with a small amount of headroom for changes in beam voltage. In pulsed operation only enough beam current is supplied to saturate at 100% power during sync. The amount of beam current reduction during video is dependent upon operating channel, perveance, available drive power and precorrection capability.

The optimum value of beam current must be experimentally determined. However, figures of merit of 77% have been achieved. The absolute value of the pulsed efficiency is directly proportional to, and therefore limited by, the efficiency obtained for nonpulsed operation. In order to prevent severe burst and back-porch distortion, a small amount of amplifier headroom is needed, approximately 2% to 5%.

Pulsed operation for klystrons is accomplished by connecting the beam control electrode to a voltage source of 0 to − 1,400 V with respect to the cathode. During sync, the pulser operates at 0 V. During the video portion of the signal, values of − 400 to − 800 V are used to achieve reliable high efficiency operation.

In some transmitters, sync is actually reduced or removed from the input video. As the beam current is pulsed, the klystron gain change increases the RF power to produce the proper sync level.

Effect of Pulsing on Transmitter Precorrection

Very little can be done to reduce ICPM by klystron tuning or selection of magnet current. Also, ICPM increases rapidly near saturation of the klystron.

The phase shift through a klystron changes due to the beam current (when the klystron is pulsed during sync). When the klystron switches back to the video current level, the previous value of phase returns. ICPM correctors operating at the exciter IF introduce a correction signal equal and opposite to the distortion produced by the klystron. A phase modulation stage in the exciter may be keyed by sync and adjusted to precorrect for incidental phase distortions caused in the klystron during pulsing.

When amplifiers are operating very close to saturation during the color burst, more differential gain and differential phase correction may be needed. Modern exciters can fully precorrect these conditions.

Sync pulse oscillations may appear as ringing when the klystron is operated at saturation. This may exhibit itself as a tearing of the picture. This ringing is believed to be caused by secondary electron feedback enhanced by the reverse gain of the klystron cavities. Rebiasing the tube slightly out of saturation will eliminate the ringing.

Multi-stage Depressed Collector (MSDC) Klystrons

Klystron amplifiers operate by converting energy from a beam of electrons to RF output power. At full output power, about half of the beam power is converted to output power. Correspondingly, half the dc input power remains on the beam as it exits the cavity region. In a standard klystron this *spent electron beam* energy is dissipated as heat. The MSDC klystron, however, operates on the *spent electron beam*, recovering its energy to reduce the dissipation. In the depressed collector technology, power recovery in the collector region is accomplished by decelerating the electrons in the spent beam. This is done by providing an electric field in the collector such that the electrons are slowed before they strike the collector wall. A collector composed of multiple elements is utilized, with each element operated at a negative potential with respect to ground. Consequently, the collector potentials are referred to as being *depressed* below ground.

The interaction process has produced a wide range of velocities for individual electrons. The collector geometry is carefully selected to provide an electric field shape which sorts the impinging electrons according to their velocity, reducing their energy as much as possible, yet ensuring that all electrons strike one of the elements and are not reflected into the electron beam.

Figure 6.2-55 shows the resulting collector configuration. The collector is composed of five elements and is designed to operate with equal voltages between elements. For a 60 kW klystron, the voltage per collector stage is typically 6.25 kV. This means that element 1 is at ground potential, element 2 is at − 6.25 kV, element 3 is at − 12.5 kV, element 4 is at − 18.75 kV and element 5 is at full cathode potential of − 25 kV. To make sure that the electron beam is optimized before entering the collector region, a refocus coil is provided just ahead of the collector.

The significant benefit of the depressed collector klystron is the reduction in power consumption for a given power output. The individual collector beam

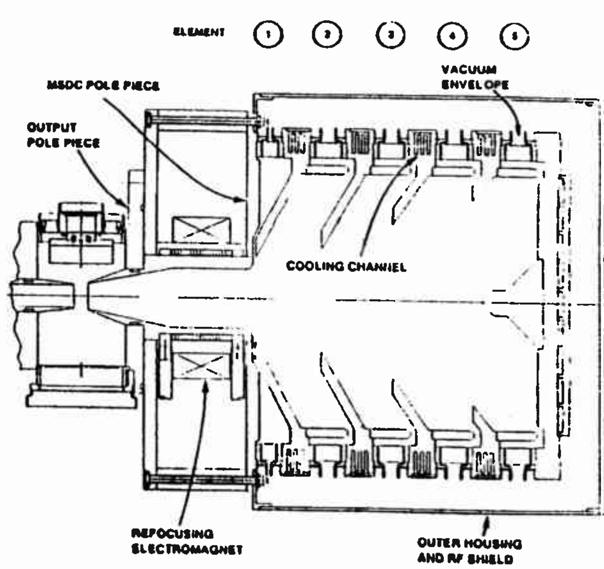


Figure 6.2-55. UHF-TV MSDC design.

currents vary depending on the picture level. Since the recovering of the spent beam takes place in the collector, it has essentially no impact on the tuning and normal precorrection of the klystron. The depressed collector technique can be applied to either external or integral cavity klystrons.

Since the power dissipated in the collector is reduced, only six gallons per minute of high purity water coolant is needed. The efficiency of the MSDC is dependent upon the transmitter configuration. Figures of merit of 1.2 to 1.6 have been obtained in transmitter installations in the field.

Transmitter Design Using Multiple Depressed Collector Klystrons

The primary differences in this transmitter from a standard klystron transmitter are in the beam supply and the cooling system. Figure 6.2-56 shows the power connections made to a MSDC klystron. Since the RF performance of the MSDC klystron is the same as the standard klystron, there are no differences in the RF driver chain.

The cooling system of a MSDC klystron transmitter uses a two-stage heat transfer system as shown in Figure 6.2-57. A two-stage system is chosen to allow outside heat exchangers to be used. The cooling system consists of a high purity water loop and a glycol-water mixture loop. High purity water (resistivity of 200,000 Ω cm or more) must be used with the cooling system to prevent current flow between collectors. To remove ions, free oxygen and other possible contaminants from the water, a three-stage purification loop is used. The filter cartridges sample part of the water flow so the whole system is continuously cleaned. The filter cartridges can be replaced without taking the transmitter off the air.

Separate beam supplies for each klystron are frequently used. Since the currents from the beam supplies change dependent on the power level, there are video frequency currents present on the power supply leads. Therefore, sufficient bypassing and power supply high voltage wire shielding is required. Monitoring each section of the beam supply current is required to obtain the currents for power dissipation calculations.

Protection circuitry should include magnet overcurrent and undercurrent trip points, beam supply overcurrent and overvoltage sensors (for each collector), water flow sensors, arc detectors within the 3rd and 4th cavities and sufficient interlocks to prevent personnel from accidentally coming in contact with high voltage.

Inductive Output Tube (IOT)

The IOT combines features of a tetrode and a klystron. Hence, the tradename *Klystrode*® is also used. The electron beam is constrained similar to that of a klystron—that is, by electromagnets. The modes of operation of the IOT are similar to that of a tetrode. However there are significant differences due to the different geometries. (Klystrode® is a registered trade mark of CPI.)

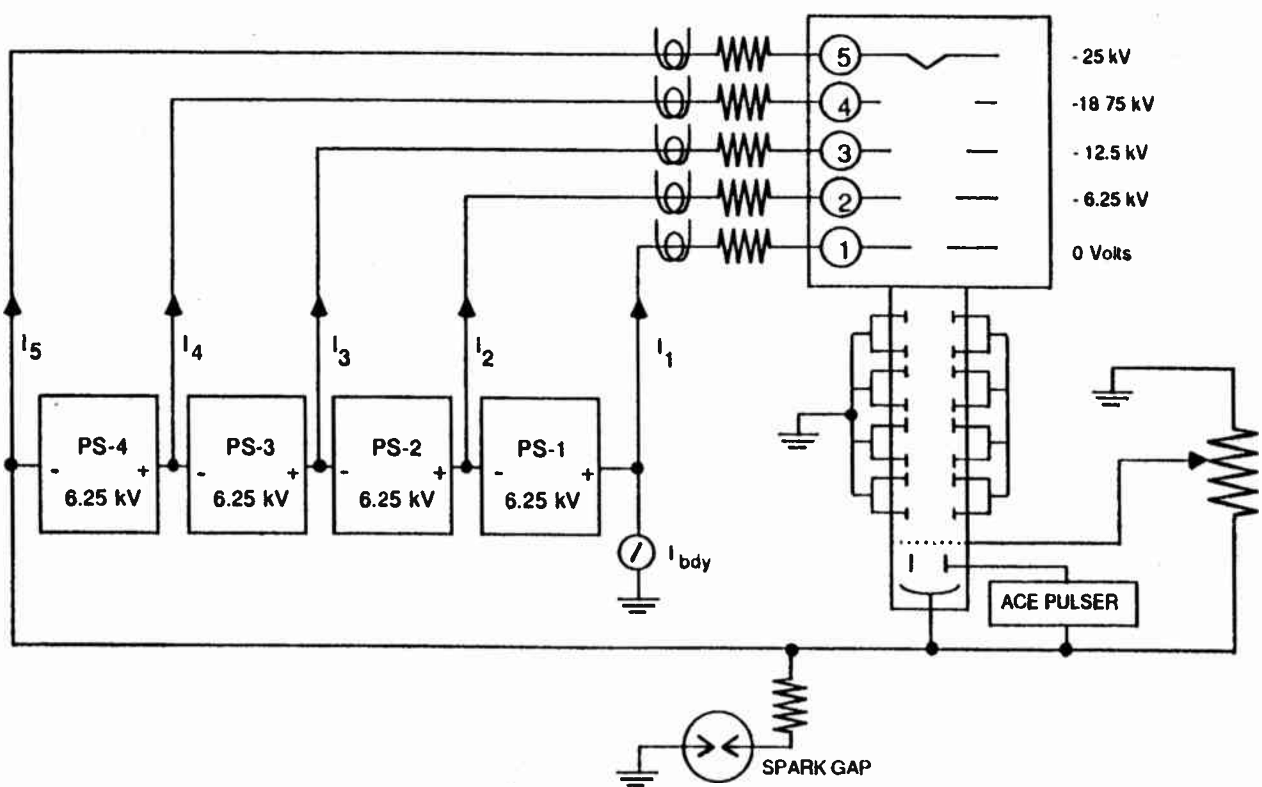


Figure 6.2-56. MSDC high voltage connections.

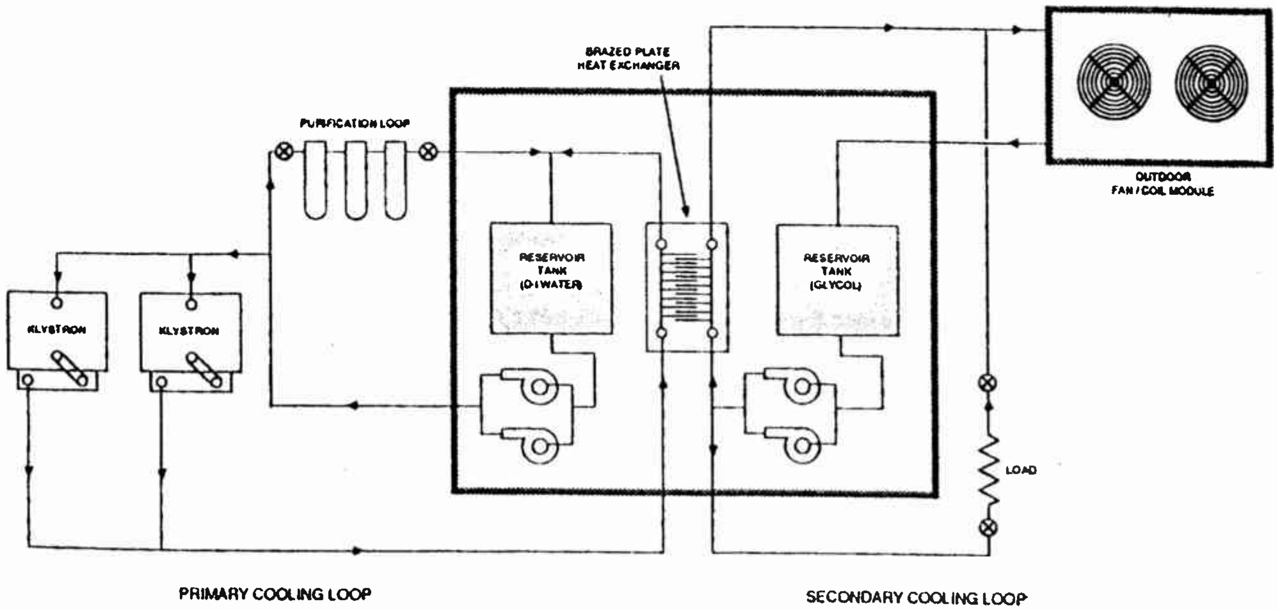


Figure 6.2-57. TV 60 UM heat exchanger system.

An IOT is shown schematically in Figure 6.2-58. It is composed of an electron gun—very similar to a klystron—a control grid, an input cavity, accelerating anode, drift tube, output cavity and collector. The electrodes are arranged linearly, unlike the concentric configuration of a tetrode. It is physically smaller than a klystron weighing approximately 400 lbs including the tuning cavities and magnet assembly. The tube itself weighs about 60 lbs.

The electron beam is formed at the cathode, density modulated with the input signal applied to a grid and then accelerated through the anode aperture. In its bunched form, the beam drifts through a field-free region and then interacts with the RF field at the drift tube gap in the output cavity. Power is extracted from the beam in the output cavity in the same manner as a klystron. The grid structure may intercept some electrons causing a small amount of grid current. This will increase if the tube is overdriven.

Input power is applied to the control grid via a resonant cavity. The grid is biased negatively near cutoff as a Class B or Class AB amplifier. The first part of the tube may be thought of as a triode with a perforated anode through which the electron beam is guided by electric and magnetic fields. The beam is bunched at the radio frequency and is accelerated by the high anode potential. It passes through the anode extension cylinder, which is an electrostatic shield, and then interacts with the RF field in the output gap. The spent beam is dissipated in the collector, which is separate from the output RF interaction circuit.

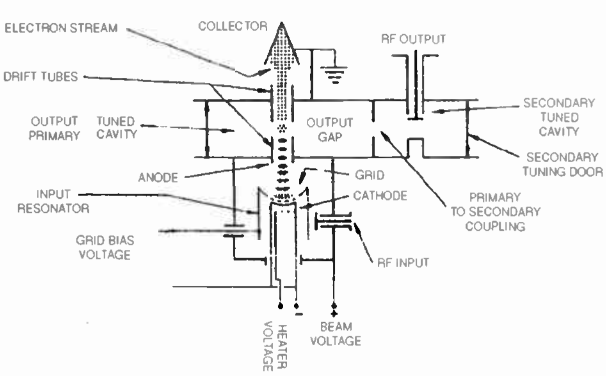


Figure 6.2-58. Klystrode schematic.

The tuned input and output circuits are external to the vacuum envelope. The input circuit consists of a stub tuner and resonant coaxial cavity, which match the drive source to the high impedance grid. Also included is a circuit which provides a dc block for high voltage. The output cavity is clamped to the body of the tube with techniques similar to those used for external cavity klystrons. The output circuit is double-tuned to achieve the bandwidth required for visual service. The double-tuned output circuit consists of a primary cavity clamped to the body and an iris- or loop-coupled secondary cavity. The output transmission line is probe coupled to the electric field or loop coupled to the magnetic field in the secondary cavity. (Coupling methods are different for different manufacturers.) Variable controls are used to adjust primary and secondary cavities to the same resonant frequency. Additional controls are used to adjust the intercavity and output couplings. Because the tube has only two cavities, it is much shorter than a klystron. The magnetic field requirements of the tube are produced by a solenoid powered by a supply of about 6 to 7 V at 25 to 30 A.

The high voltage circuitry is contained within a shielded compartment on top of the input circuit. This circuitry consists of various filters to contain the fields and prevent instabilities at video frequencies. High voltage grid bias and filament power enter this section via high voltage cables.

The fundamental benefit of the IOT is that it may operate as a Class B amplifier, resulting in high efficiency. Thus the beam current (I_b) is proportional to the RF drive signal (V_g), following the modulation envelope according to the three halves law:

$$I_a = K(V_g + V_a/\mu)^{3/2}$$

where μ is the amplification factor. The perveance (K) is proportional to the cathode area and inversely proportional to the square of the grid to cathode spacing. The drive voltage is not normally high enough to cause the instantaneous grid voltage to become positive. Although the efficiency of the IOT is dependent upon the transmitter configuration, figures of merit of 1.2 to 1.4 have been obtained for 60 kW visual service from transmitter installations in the field. In aural service, the IOT is tuned the same as for visual service. A single tube covers the entire UHF operating band, although two slightly different input cavities are required.

Power gain in either visual or aural service is about 21 dB. Thus drive power is about 500 W for the visual and 50 W for the aural (assuming 10% nominal aural power). A typical transfer curve is shown in Figure 6.2-59. The amount of nonlinearity is similar to a tetrode; the transfer function is "S" shaped. The nonlinearity of the beam current causes distortion of the fundamental UHF component and generation of signals at multiples of the UHF frequency. Harmonics do not excite the output cavity and therefore do not affect the quality of the transmitted signal. The linearity improves as the quiescent current is increased from zero, becoming optimum when at 8 to 9% of the maximum value. The transfer characteristic eventually saturates at high power levels, as is the case with all power amplifier devices. However, in an IOT, the grid voltage may be driven positive with respect to the cathode for short periods of time. This enables the tube to amplify the high power, short duration signal peaks present in digital TV signals, and also to transmit the high peak envelope power that occurs during common amplifier operation in an analog transmitter. The nonlinearity

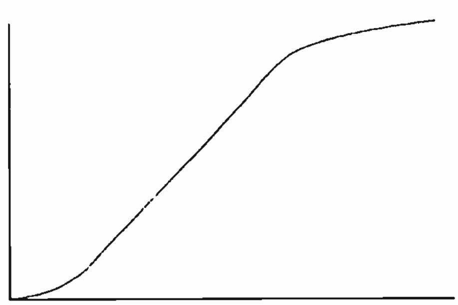


Figure 6.2-59. IOT transfer curve.

increases as the load impedance is increased to boost efficiency. Limited bandwidth of the input circuit also increases nonlinearity.

The IOT may also be used in multiplexed or common amplification mode. For example, a 60 kW tube may be used at 40 kW visual with a 10% aural signal. Intermodulation products at ±920 kHz from visual carrier generated by the tube can be precorrected by low level IF and/or RF circuitry.

Cooling of the IOT at the 60 kW power level under maximum ambient temperature conditions requires about 25 gallons per minute of water for the collector. The body of the tube is also water cooled. A 50:50 water/glycol solution is typically used in cold climates without any special water purification. The input and output cavities are air cooled. The 15 kW and 40 kW IOTs may be air cooled. The volume of air and air pressure needed are comparable to a tetrode of similar power.

Transmitter Design Using IOT

Support circuitry for the IOT consists of the necessary drivers, precorrection, power supplies and protection circuitry. An IOT transmitter block diagram using common amplification is shown in Figure 6.2-60.

The IOT uses a beam voltage of 32 kV to 36 kV. Since the beam current changes with modulation, video frequency currents are required from the beam supply. The beam supply must be designed to provide excellent regulation from no load to full load and to also provide a low source impedance for all video frequencies. In the event of a high voltage failure, the beam supply should limit the energy dumped into an arc. The supply, being stiffer, requires a triggered crowbar circuit to limit the beam supply arc energy. A block diagram of a *crowbar circuit* is shown in Figure 6.2-61.

The grid bias supply of −10 to −70 V with respect to the cathode floats with the beam voltage. A simple method of developing the grid voltage is to use zener diodes connected between the power supply and the tube cathode connection and tap the grid to the appropriate zener to obtain the desired bias current. As with klystrons, the magnet power supply must have sufficient energy storage that the beam remains focused until the beam decays. A power supply for the ion vacuum pump is needed with appropriate sensors for the protection circuitry. Other protection circuitry provided in the IOT transmitter is similar to that needed in the klystron transmitter.

TRANSMITTERS FOR DIGITAL TELEVISION (DTV)

Each of the transmitter technologies previously described have been found to be suitable for transmission of DTV. Early in the development of the DTV transmission system, klystron, MSDC and solid-state amplifiers were used for over the air transmission tests. More recently, experimental stations have gone on air using both IOT and solid-state final amplifiers. Just as for

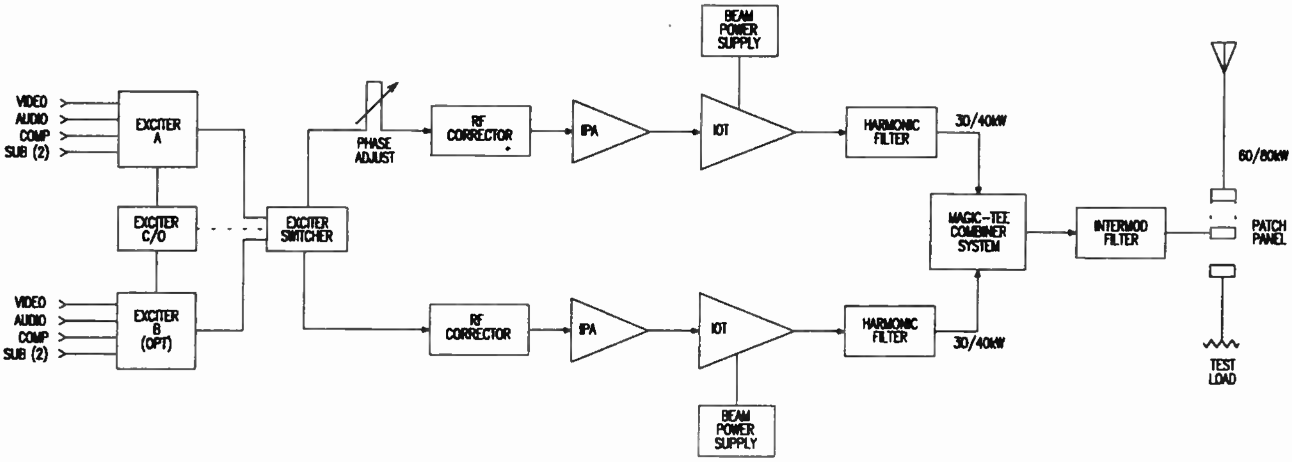


Figure 6.2-60. IOT transmitter block diagram.

NTSC, the key requirements for the final amplifier include adequate peak and average power rating, a linear transfer function and adequate bandwidth. The average DTV power afforded by a tetrode or IOT is about 40% of the NTSC peak sync power rating. Anode voltage is increased slightly from that used for NTSC for best in-band and out-of-band performance. Solid-state amplifiers will produce DTV power in excess of 30% of their NTSC peak power rating.

Even though station effective radiated power (ERP) and transmitter power output (TPO) are defined on the basis of average power, the tradition of transmitter power rating according to NTSC peak sync power has been carried forward so that most transmitters and tubes identified by means of their approximate power capability. However, average power rating is always available and should be used when selecting equipment.

Peak power rating is difficult to measure because of the random nature of the DTV signal. A peak-to-average ratio of 6 to 7 dB is usually assumed, although peaks as high as 11 dB above the average level have been measured. In practice, knowledge of the peak power is not as critical as with NTSC. The important parameters are average DTV power to support the target ERP and linearity of the transmitter sufficient to meet the FCC radiation mask specification. If these parameters are satisfied, the actual peak power need not be known with precision.

The distortion to DTV is highly dependent on the characteristics and number of driver stages. For both tube and solid-state finals, the driver stages are normally solid-state, some of which usually are operated in Class AB. The ideal transfer curve would be that of a Class A amplifier. However, this would result in high levels of dissipation and attendant cooling and power supply issues. To overcome the out-of-band products generated in the Class AB stage(s), a RF linearity correction circuit may be used. Both compression and crossover correction capability is required to suppress out-of-band intermodulation products at the band edge up to 10 dB. Alternatively, a feed forward amplifier design maybe used to provide the linearity of a Class A amplifier without the efficiency penalty. A bandpass filter is required for out-of-band suppression 3 MHz above and below the channel band edge.

The prime power required for digital transmitters is somewhat simpler to determine, compared to NTSC transmitters. The ac power required will be very nearly constant since the RF output is constant. There is no separate aural amplifier, so there is no need to consider variations in visual power and the aural power. In general, the power required will be less than for the

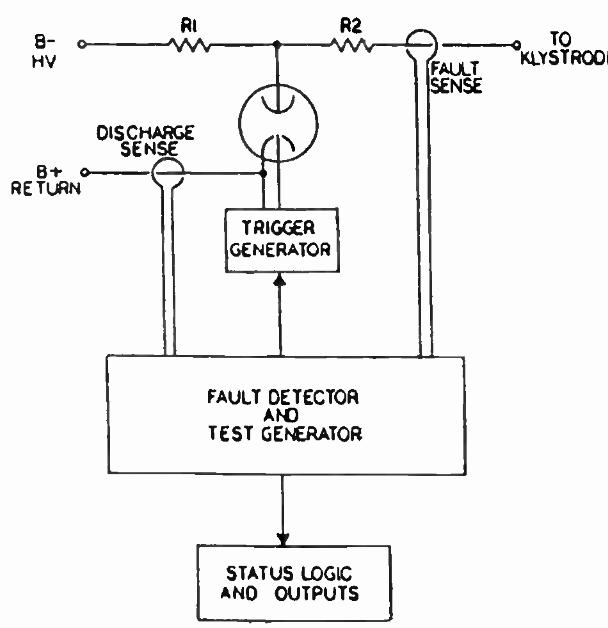


Figure 6.2-61. Triggered crowbar circuit.

corresponding peak NTSC power using the same type of amplifier technology in the same band. Likewise the cooling load is constant and somewhat lower than for NTSC.

PERFORMANCE MEASUREMENTS

The quality of the broadcast television transmitted signal is the responsibility of the broadcaster. Intermediate frequency modulation, solid-state SAW vestigial filters and sophisticated precorrection circuits make transparent performance possible. Precision demodulators with synchronous detection and SAW filters are capable of near ideal detection. Test waveforms and digital signal synthesis provide accurate test signal generation to complement transmission facilities.

The Electronic Industries Association Standard RS-508 takes into account empirical quality factors and reflects a common denominator for new transmitter performance. This standard is a valuable reference document which describes performance parameters, standards and methods of measurements.

A thorough proof-of-performance at the time of installation is an invaluable record of normal operating performance. It also serves as verification of proper signal quality and emission standards. The number of detailed measurements required after the proof usually can be limited since several performance characteristics are a measure of the same impairment. Also, some test waveforms are more useful for transmitter adjustment than others.

Transmitter measurements can be considered in two primary categories: frequency response and linearity.

In the time domain, 2T modulated sine squared pulses and multiburst waveforms are used in identifying and correcting linear distortions. In the frequency domain, swept amplitude and group delay measurements can be used, but are normally reserved for out-of-service testing.

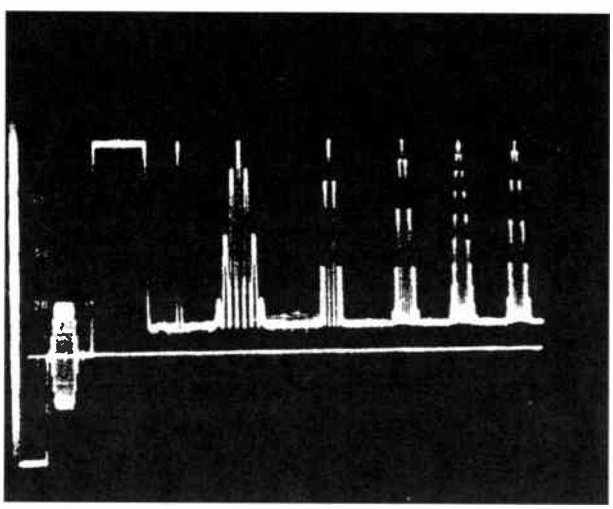


Figure 6.2-62. Multipulse test waveform.

The multipulse signal shown in Figure 6.2-62 may also be used to analyze group delay. The multipulse signal consists of a gray flag (80 IRE), 2T pulse and five sine squared pulses (one 25T pulse and four 12.5T pulses) modulated with five discrete frequencies.

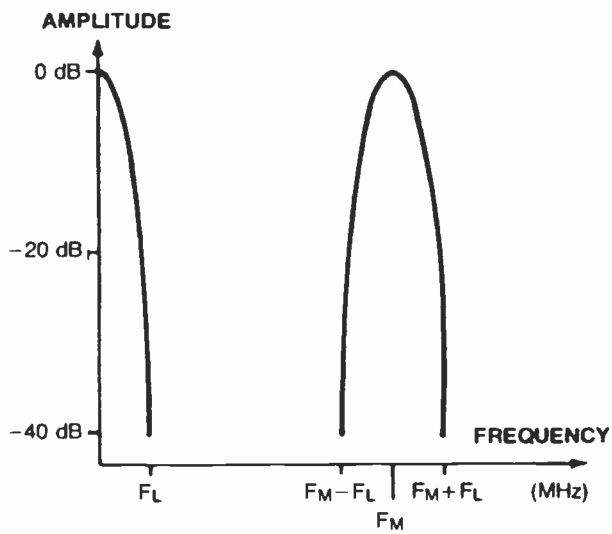
The modulated pulses contain low-frequency and high-frequency information, as illustrated in Figure 6.2-63. If the low-frequency spectra and the high-frequency spectra are delayed equally, the results will be a symmetrical modulated pulse with the same shape as the input. The gain versus frequency distortion will alter the base line flatness but will not change pulse symmetry. Delay errors will result in an asymmetrical pulse baseline. Combinations of group delay and gain errors are shown in Figure 6.2-64.

A method of quantifying these distortions uses the waveform graticule shown in Figure 6.2-65. This graticule was arrived at empirically and represents constant perceptible distortion levels. This shows that overshoots closer to the desired pulse are not as perceptible as ringing further away. The desired K factor graticule (2%) is overlayed on a waveform monitor and the group delay corrector is adjusted until the 2T waveform lies entirely within the graticule. This technique is often preferred to swept group delay measurements because the results are in terms of perceptibility.

To ascertain the carrier frequency of a station operating under precision frequency control, measurement equipment should meet the following accuracy requirements:

VHF Low Band	$+2.5 \times 10^{-9}$
VHF High Band	$+1 \times 10^{-9}$
UHF	$+3 \times 10^{-10}$

Frequency counters with the above accuracy and calibration traceable to the National Institute for Stan-



$F_L = \dfrac{1}{\text{H.A.D.}}$

F_M = MODULATING FREQUENCY

Figure 6.2-63. Energy spectrum for modulated pulses.

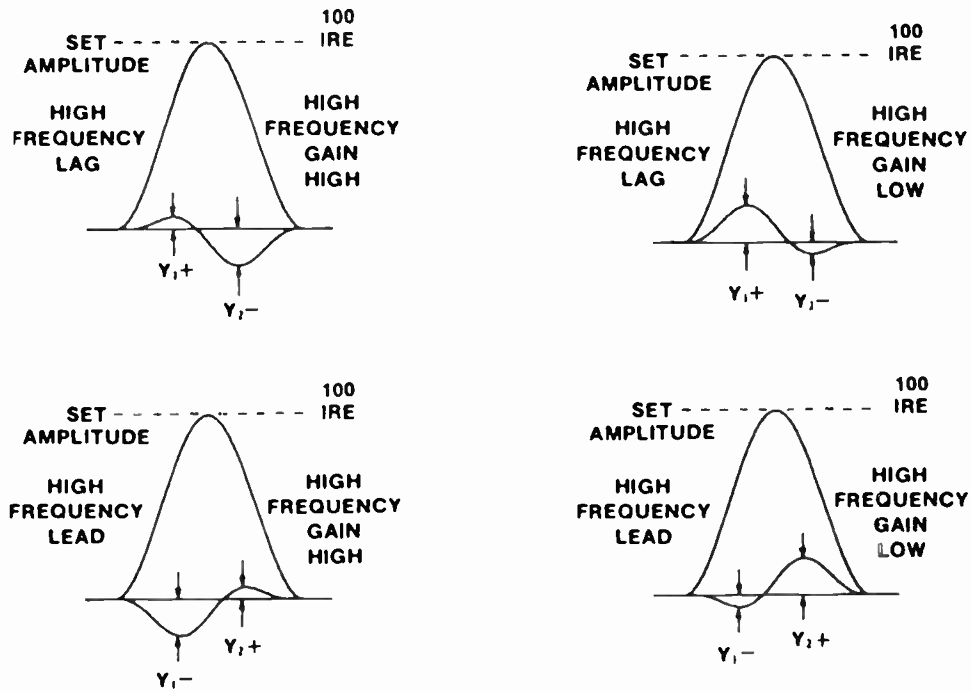


Figure 6.2-64. Modulated sine-squared pulses with gain and phase errors.

dards and Technology (NIST) may be used for making these measurements.

To test for nonlinear distortions as a function of average picture level (APL), one line of the video waveform is alternated with four lines containing a static luminance level. Figure 6.2-66 contains a modulated staircase with three average picture levels.

Modern transmitters are designed for unattended operation for extended periods of time. A properly adjusted transmitter operating with adequate cooling and regulated power lines may need only be checked in detail every three or four months or whenever a major component is replaced or repaired.

The following pretest checklist and sequence of tests are presented as a general guide for a properly adjusted transmitter. Many adjustments are interrelated and require returning to previous tests to verify proper performance. The test sequence is intended to minimize the number of adjustments.

Pretest Checks:
- Test equipment
- Input video
- Output transmission line, station load, and antenna
- Alternating current mains input
- Record meter readings, adjustment settings and key performance parameters

Transmitter Test Sequence
1. Exciter linear and nonlinear performance
2. Intermediate and driver linear amplifier performance
3. Swept frequency response

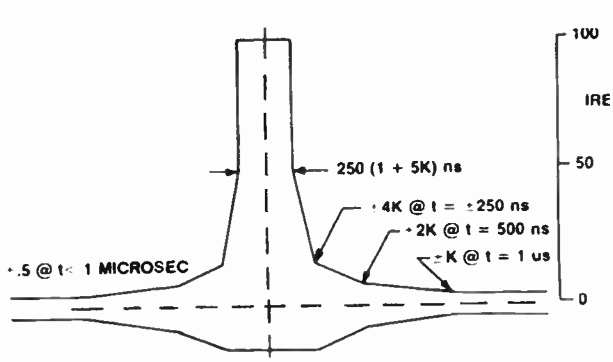


Figure 6.2-65. 2T sine-squared pulse graticule.

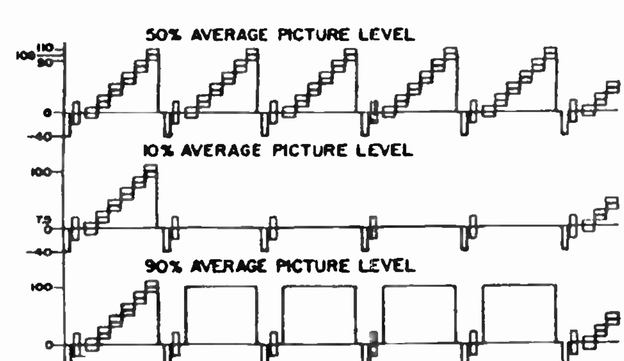


Figure 6.2-66. Modulated staircase waveform.

4. Modulation depth (includes UHF pulser and sync ICPM adjustments)
5. Power output meter calibration
6. Nonlinearity response (differential gain, ICPM, differential phase)
7. Linear response checks (group delay, pulse tests, composite waveform K factors)
8. Record meter readings, adjustment settings and key performance parameters.

During periodic measurement and adjustment, it is a good practice to record meter readings and adjustment settings before and after the test. These recordings will indicate normal setting range and can aid in getting the transmitter to near normal in the event of mistuning errors during adjustments.

In order to make satisfactory measurements of video signals, adequate test equipment is required. Measuring the output of a television transmitter at baseband requires a precision television demodulator. For greatest accuracy, this demodulator should provide a zero carrier reference pulse for determining percentage of modulation and have both synchronous and envelope detection modes.

For most baseband measurements, a video waveform monitor is all that is needed. The waveform monitor should include a provision for making measurements on vertical interval test signals (VITS) and provide filtering to allow separate examination of the luminance and chrominance components of color signals.

In addition to the waveform monitor, a vectorscope must be used for making differential gain and phase measurements. If a greater level of accuracy is desired in timing measurements, a conventional oscilloscope with an associated digital counter/timer is a great asset. For off-air monitoring, a picture monitor is convenient for quickly identifying gross video degradation.

Maintenance logs should include date, time, duration of outages, corrective action taken, and if possible, cause of failures.

Monitoring TV Multichannel Sound

Maintaining a high quality aural signal requires high quality monitoring facilities. To be able to perform the proof-of-performance measurements is another important reason to have a monitor that demodulates the RF signal and separates the components in the composite MTS signal for analysis.

An ideal modulation monitor should:

- Demodulate the composite signal from the aural carrier
- Separate the components in the composite MTS signal for measurements
- Provide in service monitoring capability
- Be suitable for proof of performance measurements
- Contain a precision (BTSC) expander.

Testing for Digital Television

The key measurements required for a DTV transmitter proof-of-performance include output power, pilot frequency, intermodulation products (IMD), error vector magnitude (EVM) IMD and harmonic levels.

1. The RMS output power may be measured with with a calorimeter.
2. The pilot frequency may be measured with a frequency counter or spectrum analyzer. The results should be the frequency of the lower channel edge plus 309,440.6 Hz, ±200 Hz, unless precise frequency control is required.
3. The FCC rules in 73.622 (h) states the power level of emissions on frequencies outside the authorized channel of operation must be attenuated no less than the following amounts below the average transmitted power within the authorized channel. In the first 500 kHz from the channel edge the emissions must be attenuated no less than 47 dB. More than 6 MHz from the channel edge, emissions must be attenuated no less than 110 dB. At any frequency between 0.5 and 6 MHz from the channel edge, emissions must be attenuated no less than the value determined by the following formula:

   $$\text{Attenuation in dB} = -11.5(\Delta f + 3.6)$$

   where: Δf = frequency difference in MHz from the edge of the channel.

   This attenuation is based on a measurement bandwidth of 500 kHz. Other measurement bandwidths may be used as long as appropriate correction factors are applied. Measurements need not be made any closer to the band edge than one half of the resolution bandwidth of the measuring instrument. Emissions include sidebands, spurious emissions and radio frequency harmonics. Attenuation is to be measured at the output terminals of the transmitter (including any filters that may be employed). In the event of interference caused to any service, greater attenuation may be required.

   In practice, it is recommended that the IMD level be measured with the resolution bandwidth set for 30 kHz throughout the frequency range of interest. This results in an adjustment to the FCC formula to account for the difference in the resolution bandwidth as stated in the rule. The adjustment is computed to be 10 log (6/.5) or 10.8 dB. Under this test condition the measured level should be −36 dB from the mid band levels to the shoulder levels.
4. Output harmonics may be measured with a spectrum analyzer. They should be at least −110 dB below the mid-band power level.
5. Measure the error vector magnitude (EVM). At the time of this writing, the only known test equipment available for measuring this parameter is a vector signal analyzer.

The output power, pilot frequency, in-band frequency response and adjacent channel spectrum should be measured periodically to assure proper transmitter operation. These parameters can be measured while the transmitter is in service with normal programming.

PREVENTATIVE MAINTENANCE

A good preventative maintenance program includes periodic inspection and cleaning of the equipment. A vacuum cleaner is preferred to remove dust instead of compressed air which will simply blow the dirt onto something else. A paintbrush can be used to dislodge dust from delicate circuit boards. Avoid using a nylon bristled brush with a plastic handle as the static charge may damage CMOS or other static sensitive components. A natural bristle brush with wooden handle and metal binding is recommended.

High voltage wires and insulators must be cleaned with denatured alcohol, or other cleaner capable of removing the dirt without leaving any residue. Meter cases are cleaned with nonstatic cleaner.

Air filters should be replaced or cleaned to maintain adequate air flow to the equipment. A second set of washable filters saves time when using a single transmitter in critical service by quickly switching the clean filters and washing the dirty ones later. Blowers should be inspected to see if the curved fins are filled with debris that would reduce air flow. Motor windings may collect a layer of dirt and interfere with the cooling of the motor itself. The fins of high-power tubes must be cleaned of debris which may have passed through the air filters. Bearings should be lubricated and checked for excessive noise.

Color change in silver plated cavity parts is a sign of overheating and may require the disassembly of the cavity to check for obstructed air passages of loose connections. Set-screws in gear and chain drive tuning mechanisms should be checked for tightness. Black silver-oxide is a good conductor and need not be removed. Small parts should be dipped in Tarnix® for cleaning. Be sure to flush the parts after cleaning to remove any residue. Scotch-Brite® is a good nonmetallic cleaning pad for silver plated parts. Do no remove the plating when cleaning these parts.

High current wires may move during turn-on surges and can suffer abrasions which may cause an arc if not properly dressed from sharp edges. Wiring on terminal boards may loosen through thermal cycling and vibration. All connections must be checked to be sure they remain tight. If wires need to be replaced, the correct gauge, voltage rating and temperature rating must be considered when selecting the replacement.

Edge connectors on printed circuit boards should be cleaned with Cramolin® or other cleaner. A small amount is applied to the connector and then removed with a lint-free cloth. Do not use pencil erasers as this will remove gold or silver plating from the edge traces, and could degrade the connection or create an intermittent. The sulfur in the eraser causes chemical reaction to the edge connector material.

Backup systems or emergency modes of operation should be checked periodically. Relays should be exercised to keep contacts polished.

Transmitter site cleaning should include a check of the building for such things as leaks in the roof which may cause damage to the transmitter, and the presence of insects or small animals.

Intake blowers with filters capable of creating positive pressure in the room can minimize the need for cleaning.

Careful records must be maintained in order to establish a good history for future reference. Such a log will include a description of what was done, when it was done and the name of the person that performed the work. Seasonal events such as harvesting in farm locations, severe weather, construction projects in and around the transmitter building can require special action, but usually a pattern will emerge that allows the maintenance to be scheduled on a regular basis.

A complete set of meter readings taken when the equipment is working properly and updated weekly or monthly can greatly assist problem diagnosis.

Create a maintenance program with weekly, monthly quarterly, semi-annual and annual tasks evenly spread throughout the year. The following list of items can be used to develop a maintenance routine:

Maintenance Items

- Prefilter manometer readings
- Inspect prefilters
- Replace prefilters
- Post-filter manometer readings
- Cabinet input air manometer reading
- Inspect transmitter air filters
- Replace transmitter air filters
- Vacuum cabinets
- Clear tube fins
- Measure blower currents
- Clean fan blades and motor windings
- Check connections for tightness
- Inspect MOVs.

Recommended Data to be Recorded

- All parameters on meters or user displays
- Direct current input power and calculate dissipation
- Transmitter currents in black picture and at idle (no RF drive).

Visual Performance Checks

- Ensure proper video level
- Establish proper sync level
- Optimize differential gain
- Optimize ICPM
- Optimize differential phase
- Optimize group delay using T pulses
- Assure proper power calibration
- Secure proper audio processor setup
- Optimize swept response.

Aural Performance Checks

- Ensure proper modulation levels
- Establish proper SCA input levels
- Assure proper power calibration

- Secure proper audio processor setup
- Optimize audio frequency response
- Check and minimize distortion
- Optimize stereo separation
- Check and minimize crosstalk between MTS channels.

Control System Checks

- Authenticate proper operation of all interlock circuits
- Verify proper operation of all overload circuits
- Confirm proper operation of all control processes (VSWR foldback, filament timing, coax switches, etc.).

AIR SYSTEMS FOR TRANSMITTERS

Transmitters dissipate large amounts of power. Air which has already passed through the equipment and has picked up heat must be removed from the immediate vicinity to prevent the hot air from being recirculated. In addition, provisions must be made for sufficient intake air to replace that which has been circulated.

To provide adequately for hot air exhaust and fresh air intake, the maximum and minimum environmental conditions in which the equipment may operate, and the following information from the equipment manufacturer should be known:

Altitude: ______________________ (feet)

Max. temp: ______________________(°C)

Min. temp: ______________________ (°C)

Total air volume through the transmitter: ________ ____________ (CFM)

Pressure drop within the critical portion of the air circuit (across the transmitter tube for example) ____________ (inches of water)

Air temperature rise through the transmitter: ______ ____________ (°C)

Air exhaust area (A): ________________ (Sq. Ft)

Equipment layouts usually provide an area for heated air to exit from the top of the cabinet. The size and location of this exhaust area is usually shown on a manufacturer supplied outline drawing.

Most broadcast equipment internal air systems are designed to be operated into free space (back pressure of 0.0 in. of water) so exhaust ducts must have minimum loss. Good practice is to design for no more than +0.1 in. of water pressure in the duct close to the exhaust area of the transmitter.

Any exhaust installation other than a large cross-section duct (equal to the cross-section of the transmitter exhaust port) with no bends, and with a long radius turn outside the transmitter building for weather protection, will need an exhaust blower or fan.

The recommended system is sized only for cooling the transmitter. Any additional cooling load in the building must be considered separately when selecting the air system components. The transmitter exhaust should not be the only exhaust in the room as heat from the peripheral equipment would be forced to go out through the transmitter.

The *sensible-heat* load is the sum of heat loads such as solar radiation, heat gains from equipment and lights and personnel in the area that is to be cooled.

The following exhaust duct design illustrates the key cooling concepts:

$$\text{Air volume through transmitter} = 325 \text{ CFM (ft}^3\text{/min)}$$

$$\text{Air exhaust area (A)} = 3.4 \text{ square ft}$$

$$\text{Air exhaust velocity (V)} = \frac{325 \text{ CFM}}{3.4 \text{ ft}^2} = 94.5 \text{ ft/min}$$

Air velocity of 94.5 ft/min is relatively low allowing a transition to a smaller diameter pipe if desired. Assume a transition to a 10 in. diameter pipe.

$$A = \frac{\pi D^2}{4 \times 144} = \frac{\pi(100)^2}{4 \times 144} = .545 \text{ square ft}$$

$$V = \frac{325 \text{ CFM}}{\text{dia. pipe } .545 \text{ ft}} = 596 \text{ ft/min}$$

The air friction in a 10 in. diameter pipe is 0.06″ per 100 ft of pipe at a flow of 325 CFM (see Figure 6.2-67). Assuming there is 20 ft of straight pipe to a roof, with two 90° elbows to turn the pipe down for weather protection, the total loss of the exhaust is estimated as follows:

$$\frac{20}{100} \times 0.06'' = 0.012 \text{ inches of water (pressure drop in 20 ft of pipe)}$$

Consider the static pressure drop of two 90° elbows, to give a 180° bend in the pipe. From Figure 6.2-68

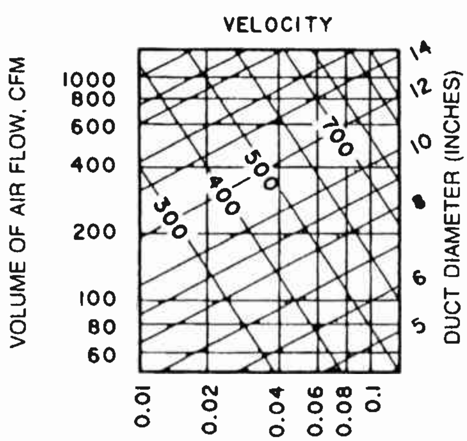


Figure 6.2-67. Friction loss in pipes.

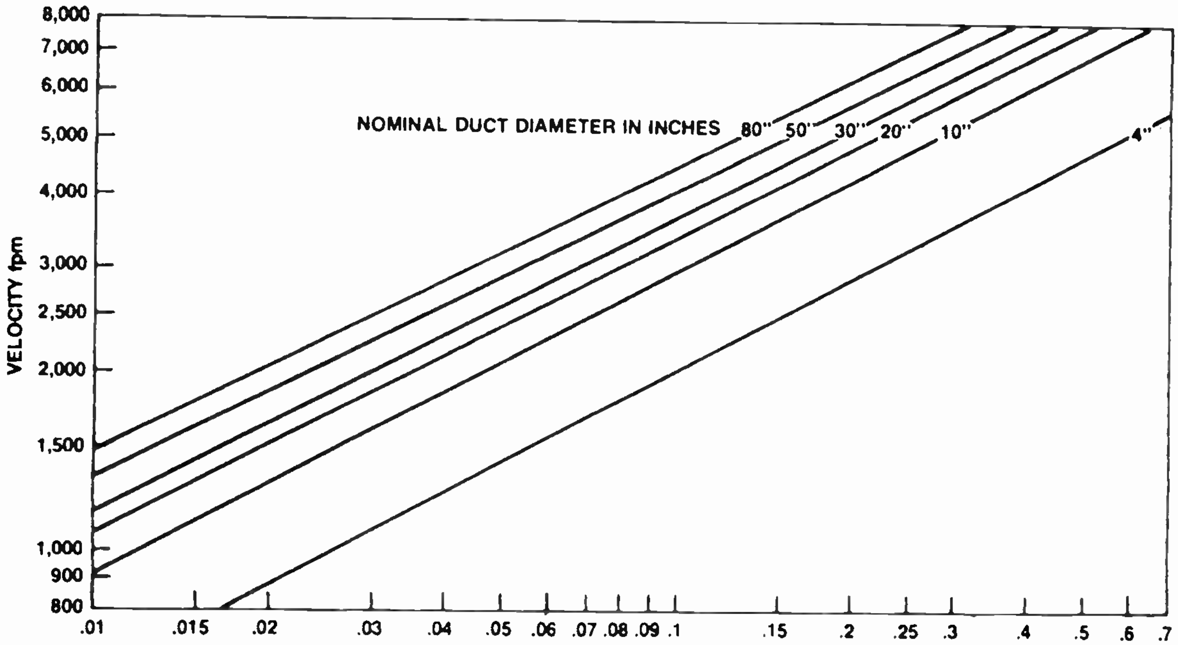


Figure 6.2-68. Friction loss in 90° elbow.

a 10 in. elbow at 900 ft/min gives less than 0.01 pressure drop. There is just under 600 ft per minute in the system. Conservatively, add .01 for each 90° elbow and .12 for the 20 ft section of 10 in. pipe.

0.012″ + 0.01″ + 0.01″ = 0.032 inch of water (total pressure drop)

No exhaust fan is necessary since 0.032 in. is less than the 0.1 in. pressure that requires a fan.

The installer may have a problem exhausting the transmitter in this simple fashion (for example a roof exit is not available). It may be required to add two additional 90° elbows and a straight run of 10 in. pipe 100 ft long. The pressure drop in the exhaust system is then:

Friction loss in 20 ft of 10 in. dia. pipe	0.012
4 elbows × .01 in. of water (each)	0.04 in.
Friction loss/100 ft of 10 in. dia. pipe	0.06 in.
	0.112 in. of water

The 0.112 in. of pressure now exceeds the level at which a fan is needed.

The performance curve for the fan, shown in Figure 6.2-69, indicates that it will deliver 325 to 330 CFM into 0.1 in. of water. This is sufficient to handle all transmitter air and overcome all estimated duct losses.

The outline drawing in Figure 6.2-70 shows a typical exhaust duct and blower system. The recommended minimum ceiling height to properly handle exhaust air as shown is 12 ft. The outline drawing also indicates a typical air intake and prefilter system.

These calculations are for a very simple air system. Most television transmitters have multiple enclosures with large volumes of air required. In these cases, the equipment user is required to combine the heat exhaust air from several enclosures into a complex duct system. When these circumstances occur, the safest practice is for the user to contract for the services of a heating, ventilating and air conditioning (HVAC) consultant. The cost involved in having the best possible air ex-

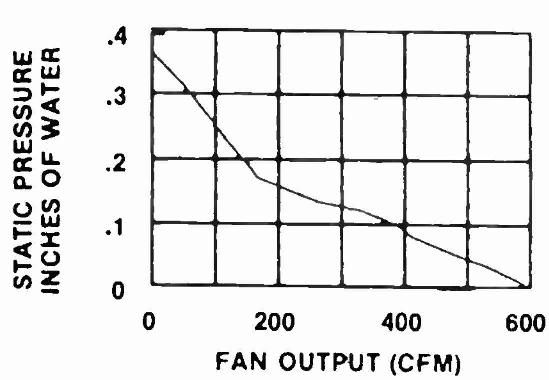


Figure 6.2-69. Fan performance curve.

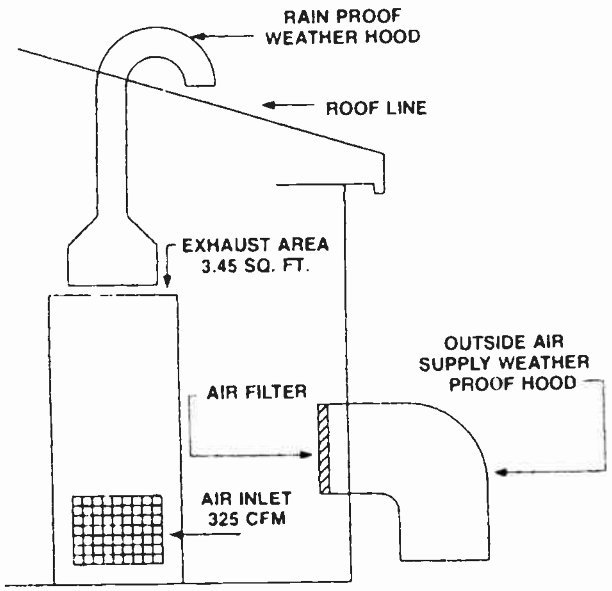


Figure 6.2-70. Suggested intake and exhaust duct installations.

haust and supply system will pay in extended transmitter life and lower service costs.

The intake vent and blower should be sized to provide a slight positive room pressure. Installing a manometer to sense pressure drop across the filters helps determine replacement interval of prefilters.

If existing space will not permit the construction of the transmitter manufacturer's recommended air system, care must be taken to modify the design to fit the available space and still properly cool the transmitter.

Additional flushing air is recommended for the removal of heat from any surrounding equipment that shares space with the transmitter. It is a recommended input air be no greater than 5°C above ambient.

Air Conditioning

It is a common practice to set the transmitter into a sealed wall forming a plenum chamber. The plenum is supplied with outside air and separate air conditioning is provided for the front side to cool personnel and source equipment.

In areas with severely polluted air, it may be necessary to run the transmitter on air conditioned air to avoid bringing in corrosive salts or gaseous contaminants.

The amount of air conditioning will depend on several factors. It is strongly recommended that the air conditioning be shared by a number of units rather than one large central system so that operation can continue in the event of the failure of one unit. Air conditioning units are usually specified by *ton* of cooling capacity with *one ton* equal to 12,000 btu per hour.

Consult experienced professionals in the area of HVAC design to achieve the desired result and to prevent problems.

KEY TERMS

Baseband. The video, audio or digital signal supplied from program input equipment. For analog TV in the United States, these signals are defined by the NTSC standard. For digital TV in the United States, the input signal is defined by the Advanced TV Standards Committee Standard A53.

Efficiency. Fundementally, the ratio of RF output power to input power, expressed as percent. For TV transmitters, special conditions apply (see section on TV transmitter efficiency).

Reliability. The degree to which equipment will operate without failure or off-air time. Quantitatively, reliability is often expressed as the *mean time between failures* (MTBF); alternatively, availablity is a useful measure.

Linearity. Refers to the degree with which the transmitter output voltage is directly proportional to the input. Common terms used to quantify the degree of transmitter nonlinearity include low-frequency or *luminance nonlinearity*, *differential gain* and *AM to AM conversion*. Output phase may also be a function of input level. The deviation from ideal linear phase is often quantified as incidental carrier phase modulation (ICPM), differential phase and AM to PM conversion.

Differential Gain. Nonlinear chroma gain as a function of luminance level. A change in the color saturation results.

Differential Phase. Nonlinear chroma phase as a function of luminance level. A change in hue results.

Low-Frequency or Luminance Nonlinearity. The change in luminance gain as a function of brightness level.

Linear Distortions. Distortions to the transmitted signal that are not level dependent. They are introduced by linear components in the transmission path. These components include any device with a non-constant frequency response such as matching networks, cavities, filters, diplexers and other tuned circuits. Variations in amplitude and phase are included.

Group or Envelope Delay. Nonuniform delay of different frequencies over the signal bandwidth; the first derivitive of phase with respect to frequency. It is caused by nonlinear phase as a function of frequency inherent in RF amplifiers, filters, combiners and other output devices. In general, the closer the amplitude roll-off is to the passband, the higher the group delay distortion.

Nonlinear Distortions. Distortions to the transmitted signal introduced by non-linear components in the transmission path. These components include any device whose output voltage is not directly proportional to input such as power amplifiers operating near compression.

Precorrection. A technique or circuit introduced to compensate for nonlinear distortion. The objective is to provide a complementary transfer function that, when operating on the nonlinear transfer function, will minimize nonlinear distortion. Precorrection may be introduced in the baseband, IF or RF sections of the system and may be manually or adaptively adjusted.

Equalization. A technique introduced to compensate for linear distortions. The objective is to provide a complementary transfer function that, when multiplied with the frequency response function, will minimize linear distortion. Equalization may be introduced in the baseband, IF or RF sections of the system and may be manually or adaptively adjusted.

Transport Layer Interface. (The interface between the DTV baseband signal generation equipment or studio to transmitter link (STL) and the DTV transmitter.) It is defined by SMPTE standard 310 M.

Luminance Nonlinearity. The change in luminance gain as a function of brightness level.

Surface Acoustic Wave. The propagation of waves on the elastic surface of a piezoelectric crystal. The wave propagation is roughly the speed of sound, and therefore called *acoustic.*

Common Amplification. Use of a power amplifier with sufficient bandwidth and linearity in which the visual and aural signals are combined at low level and amplified together.

Microphonics. The susceptibility of a circuit or system to mechanically induced phase and frequency shifts.

Error Vector Magnitude (EVM). A useful quantitative measure of the quality of a digital transmission system or subsystem. It is the RMS value of the vector magnitude difference between the ideal constellation points and the actual constellation points of the I/Q diagram, expressed in percent. EVM is usually computed as an average over several samples. A perfect digital transmission system would exhibit an EVM of 0%.

Eye Pattern. A convenient method for visually assessing the performance of a digital transmission system. The signal is displayed on a scope set to trigger at the symbol rate. The persistence of the scope creates a composite of all possible waveforms. At each level of the 8 VSB signal, the resulting pattern resembles the human eye. The degree to which the eyes are open is a measure of the signal quality. Eyes open 100% correspond to EVM of 0%.

REFERENCES

ANSI/IEEE C62.41 Standard (Sometimes referred to as the IEEE-587 Standard).

Electrical Performance Standards for Television Broadcast Transmitters, Electronics Industries Association, RS-508.

Engineering Design Manual for Air Handling Systems, United Sheet Metal Bulletin 100-5-178.

Electronic Industries Association, "BTSC System Recommended Practices" *EIA Systems Bulletin No. 5,* July 1985.

Eilers, Carl G., "The In-band Characteristics of the Vestigial Sideband Emitted Signal for ATV Terrestrial Broadcasting," *IEEE Transactions on Broadcasting,* Volume 42, December 1996.

Gysel, Ulrich, "A New N-Way Power Divider/Combiner Suitable for High Power Applications," *IEEE-MTTS-5, International Symposium Digest,* 1975.

Mark's Mechanical Engineers Handbook, McGraw-Hill, 6th Edition, Edited by Theodore Bawmeister, 1958.

Plonka, Robert, "Group Delay Corrector for Improved TV Stereo Performance," *NAB Broadcast Engineering Conference Proceedings,* 1989.

Rhodes, Charles, "The 12.5T Modulated Sine-Squared Pulse for NTSC," *IEEE Transactions on Broadcasting,* Volume 18, March 1971.

Sgrignoli, Gary, "Measuring Peak/Average Power Ratio of the HDTV Signal with Vector Signal Analyzer," *IEEE Transactions on Broadcasting,* Volume 39, June 1993.

The original information on tetrodes for UHF transmitters was supplied by Timothy P. Hulick, Ph.D. Acrodyne Industries, Inc., Blue Bell, PA.

Tew, Barry and Collins, Gerald W., *DTV Transmitters Using New Technology,* Presented at the Montreaux International Television Symposium, June 1997.

Weirather, Robert, "A Distributed Architecture for a Reliable Solid State VHF Television Transmitter Series," *NAB Broadcast Engineering Conference Proceedings,* 1989.

ADDITIONAL SOURCES OF INFORMATION

BOOKS

Weiss, Merrill S., *Issues in Advanced Television Technology*, Focal Press, 1996.

Gonzolez, Guillermo, *Microwave Transistor Amplifiers,* Prentice Hall, 1984.

Whitaker, Jerry C., *Maintaining Electronic Systems,* CRC Press, 1991.

6.3
MULTICHANNEL TELEVISION SOUND

EDMUND A. WILLIAMS
ADVANCED TELEVISION SYSTEMS ENGINEERING
PUBLIC BROADCASTING SERVICE
ALEXANDRIA, VA[1]

INTRODUCTION

Multichannel television sound (MTS) is a generic term for the process of adding subcarriers to the aural carrier of an NTSC television station. The development of a stereo MTS system was accomplished through the Broadcast Television Systems Committee (BTSC) of the Consumer Electronics Manufacturers Association (CEMA), formerly known as Electronic Industries Association (EIA), and is known as the BTSC system for television stereo sound. Subcarriers in the BTSC system are designed to be received by the public and may be used for stereophonic sound and, simultaneously, a secondary audio program (SAP) channel for a second language or other program related aural service. A third subcarrier, in the BTSC system, may be used by the broadcaster for professional uses such as news crew cues, transmitter telemetry or other digital or analog, aural or data services. The BTSC stereo system is completely compatible with monophonic receivers.

Non-BTSC subcarriers may be used for a different stereo system or for non-program related material. However, such signals must not interfere with BTSC equipped television receivers.

Modern broadcast television transmitters, built after 1984, are designed to accommodate multichannel sound. Older transmitters are likely to require revised tuning procedures and more attention to total performance will be necessary to prevent degradation to the stereo signal. Most importantly, new testing and measurement techniques are required to insure that the BTSC multichannel sound signals are transmitted to the receiver in good condition.

The introduction of BTSC multichannel sound as an all-industry recommendation allows receiver manufacturers to develop new models that offer good stereo performance. In 1997, over 50% of all television receivers sold in the U.S. had BTSC stereo. Nearly 50% of all U.S. TV homes have one or more stereo television receivers. Nearly all television programs are produced in stereo. More than 90% of all television stations in the U.S. have BTSC transmission capability. Nearly all video services available to the public (VCR, Cable TV, satellite, video disk, DVD) contain stereo sound tracks.

The adoption of the BTSC system accelerated the use of stereo sound tracks for television program production. Handling the stereo sound throughout the broadcast facility in turn created challenges to audio design and production, installation, distribution, operation and maintenance.

During the introduction of the new stereo service, many stations employed stereo synthesizers to provide viewers with an immediate sensation on newly acquired stereo television receivers. Because a large portion of original program material is now produced in stereo, the synthesizer is no longer needed, nor desired, in the audio chain. Using the synthesizer anywhere in the audio chain other than in the production studio can severely degrade mono and stereo sound (and enhancements such as surround sound) which in turn may cause distraction and discomfort to the viewer.

Background

The EIA established a laboratory and conducted transmission tests on several multichannel television sound systems. The system selected by the committee uses the transmission parameters developed by Zenith Electronics Corporation and the noise reduction system developed by dbx Incorporated. This combination, called the BTSC Multichannel Television Sound system, was recommended to the FCC as an industry-wide standard and adopted on March 29, 1984 in the *FCC's Second Report and Order of FCC Docket 21323.* While not specifying the BTSC system as such, the FCC provided protection to the establishment of the BTSC system that in turn has resulted in a well developed *de facto* standard. Some excerpts from the FCC rules include:

1. BTSC equipped receivers are protected from interference from other MTS subcarrier schemes (see Section 73.682(c) of the FCC CFR).
2. *Bulletin OET-60A*, published by The FCC Office of Science and Technology (formerly Engineering and Technology), is the document that contains the BTSC specifications described in FCC Rule 73.681.
3. Stations are permitted to implement any other MTS system of subcarriers needed if no interference is caused to receivers designed to BTSC specifications or to regular monophonic sound television receivers.

While the BTSC system is not a standard as such, the Commission provided the next best thing—protection. Therefore, transmitter and receiver manufacturers can

[1] Contributions to this chapter provided by Eric Small, Modulations Sciences, Inc., Somerset City, NJ.

produce equipment conforming to the BTSC system with full knowledge that a compatible, protected MTS system is recognized by both FCC and industry. CEMA also developed and published *BTSC System: Television Multichannel Sound Recommended Practices*, a set of recommended operating practices aimed mainly at transmitter and receiver manufacturers.

Television stations may employ aural subcarriers for virtually any purpose as long as they do not interfere with the BTSC system or normal monophonic receiver operation.

All stations contemplating stereo operation should obtain a copy of *Bulletin OET-60A* entitled *Multichannel Television Sound Transmission and* Audio *Processing Requirements for the BTSC System.*

It should be noted that the multichannel sound system adopted for use in the U.S. is unique. Multichannel sound transmission schemes used in other countries have different configurations. The Japanese MTS carries either a stereo sound channel or a second language channel but not both simultaneously. West Germany adopted a two carrier sound transmission system providing either stereo or second language. In Great Britain, a digitally modulated carrier for stereo sound is inserted just above the normal monaural FM carrier. Except for Japan, all countries that employ the NTSC system have adopted the BTSC MTS system. In comparison, the U.S. system with its combination of AM and FM subcarriers, can simultaneously transmit stereo and a separate program or second language, and a subcarrier for professional or non-public uses. Further, the U.S. system allows stations to employ other subcarrier arrangements, when not using the BTSC system, to serve specific station requirements.

A noise reduction or *companding* (compression and expansion) system for both the stereo and separate audio program channels was designed into the BTSC system from the beginning. A high performance compressor is built-in to all BTSC encoders. A matching expander is built-in to all BTSC decoders. They are, therefore, part of the complete system design and not simply an add-on feature.

THE BTSC SYSTEM

The design of the BTSC MTS considered several major performance objectives and considerations including:

- Full sound fidelity
- Compatibility with existing monophonic sound television receivers
- Ease of implementation in transmitters
- The ability of the MTS signals to pass through cable television systems, master antenna systems and TV translators
- The ruggedness to withstand typical transmission impairments (noise, multipath, interference) in the path between the transmitter and receiver.

The BTSC MTS is designed to be compatible with existing monophonic television receivers and provide high quality stereo sound transmission. The system requirements and performance capability are as follows:

- The BTSC MTS is compatible with existing monophonic television receivers. Establishing the correct operating levels in the stereo encoder is essential for proper stereo and monophonic operation. Monophonic performance is also a function of the balance of the original audio program material and the distribution and processing within the broadcast plant.
- The system provides both a stereo and a separate audio program (SAP) channel simultaneously. The SAP channel is available on most BTSC equipped television receivers. The quality of the SAP channel, while limited compared to the stereo and mono channels, depends on the original program material and the processing used in the broadcast facility.
- There are provisions for other professional use (PRO) or station use subcarriers. The PRO channel may be used for a variety of purposes for station operations or commercial services. It cannot be received on consumer television receivers.
- A noise reduction (companding) system is employed for both stereo and second audio program channels. Nearly 30 dB subjective reduction in noise is achieved through circuit design, compression algorithm, precision components, proper setup and routine maintenance. The decoders used in all consumer BTSC equipped television receivers are matched to the encoders used in the transmitter. CEMA publishes a detailed description *EIA Television Systems Bulletin #5*, of the entire BTSC system entitled *Television Multichannel Sound BTSC System Recommended Practices* which is used by manufacturers of television transmitters, receivers, cable television headend equipment and add-on adapters and other equipment that contains BTSC encoders and decoders.
- The aural FM carrier frequency deviation capability in the aural transmitter exciter and associated amplifiers must be increased to accommodate the new BTSC subcarriers. The peak deviation for monophonic transmission remains at 25 kHz. With stereo, SAP and PRO the main FM carrier peak deviation will increase to 73 kHz. Insufficient bandwidth in aural transmission components will result in distortion and crosstalk between audio channels in the receiver.

Excellent stereo audio quality and separation between stereo and SAP channels can easily be obtained from BTSC stereo exciters. However, the transmission system will degrade some performance parameters, depending upon how well the RF components are adjusted and maintained.

Table 6.3-1 shows the MTS performance objectives that can be achieved with properly operating BTSC and aural RF transmission equipment, routine maintenance and use of good quality and calibrated test equipment.

Degradation to any of the performance parameters will occur under the following conditions:

- Crosstalk: Non-linear RF amplifiers, excessive incidental carrier phase modulation (ICPM.)

Table 6.3-1
MTS performance objectives.

Channel	Frequency Response	Noise	Distortion
Main (monophonic)	20–15 kHz ± 1 dB	>58 dB	<1.0%
Stereo (difference)	100–15 kHz ± 1 dB	Dyn*	<1.0%
Separate Audio Program (SAP)	50–10 kHz ± 2 dB	Dyn*	<2.0%
Professional Use (PRO)	100–3 kHz ± 3 dB	>40 dB	<5.0%
Stereo Separation	>40 dB (without NR encoding)		
Crosstalk into SAP or PRO	>50 dB		
Crosstalk into Main (any)	>60 dB (40 dB for stereo into main)		

*Dyn indicates that the noise varies dynamically with program content and the characteristics of the noise reduction system. Subjectively, the noise will normally be 60 dB below peak operating levels.

- Poor frequency response, noise and distortion: Audio circuitry in stereo encoder or exciter.
- Noise reduction system artifacts such as pumping or changes in level: Improper adjustments in encoder (do not adjust without knowledge of the system and proper instrumentation).
- Buzz in the received audio: Excessive ICPM in the transmitter, video over-modulation, video energy in the aural carrier passband.

The BTSC System Transmission Format

The transmission format of the BTSC baseband configuration is shown in Figure 6.3-1. The monophonic component (L + R) is transmitted as a standard FM signal as in the past. The stereophonic component is transmitted as a difference or left minus right (L − R) signal and based on the current U.S. FM stereo broadcast model that is familiar to most broadcast engineers. Subcarriers for cues, data, telemetry or other station related purposes may also be employed. Space for professional use, non-public subcarriers is provided in the aural baseband. Figure 6.3-2(a) shows how the BTSC signal is configured at the transmitter. Figure 6.3-2(b) shows how BTSC is handled in a typical receiver.

Monophonic Channel

The monophonic, or sum channel (L + R), is maintained at its present place in the aural baseband and continues to have 25 kHz peak deviation and the standard 75 μsec preemphasis. The addition of BTSC or MTS subcarriers, while increasing the total deviation of the main FM carrier to 73 kHz, does not result in any degradation to the normal monophonic sound in existing receivers. The monophonic channel is maintained at 25 kHz deviation unlike FM radio broadcasting where the mono channel is reduced when transmitting stereo. This would result in lowering the signal-to-noise ratio of the aural channel and the loudness of the sum channel compared with stations not operating with the BTSC system. Increasing the total deviation provides the room for the stereo, SAP and PRO subcarriers to be added to the aural baseband.

The fidelity of the main channel monophonic audio is not adversely affected. However, because the sidebands of the stereo subcarrier are close to the main channel audio, and to prevent crosstalk between channels, BTSC stereo encoding equipment incorporates

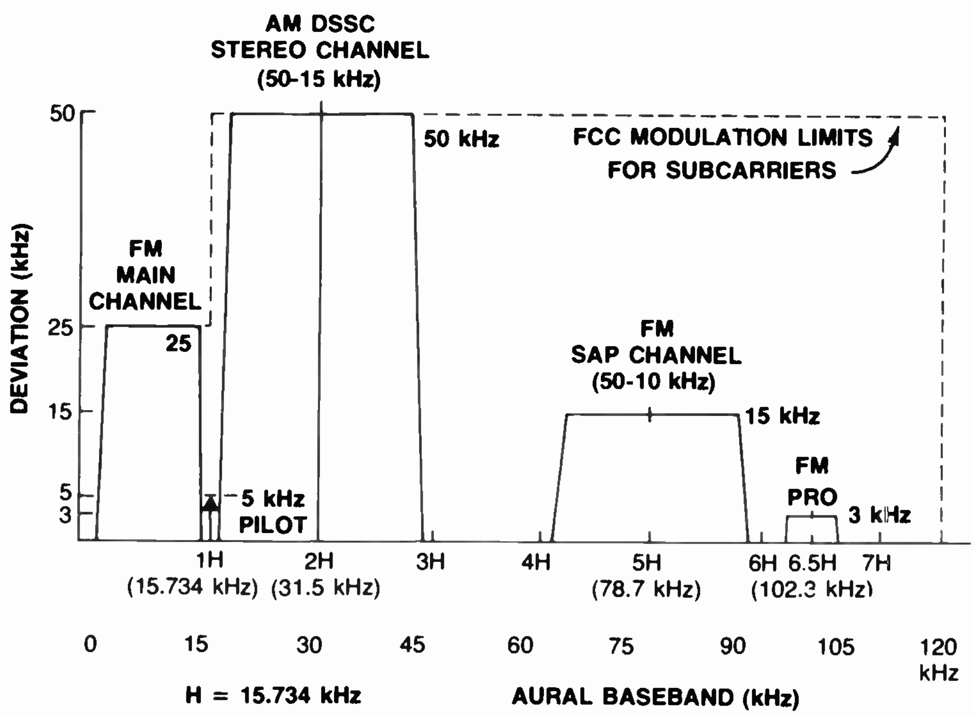


Figure 6.3-1. BTSC MTS baseband.

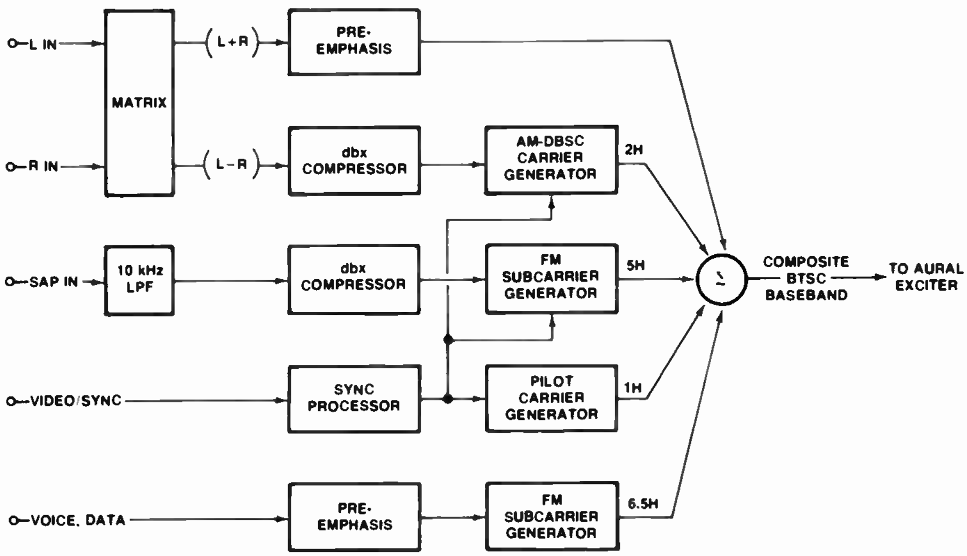


Figure 6.3-2(a). Typical MTS transmitter encoder.

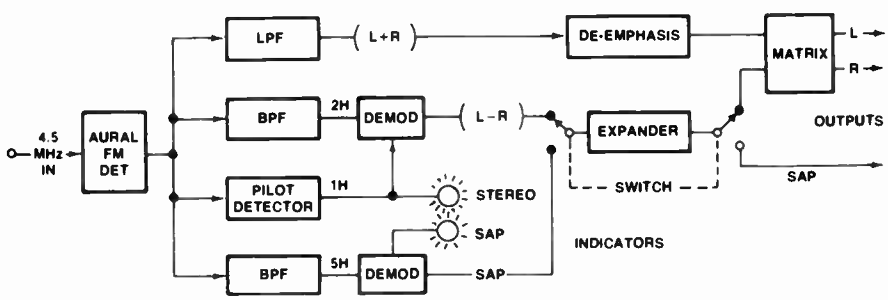


Figure 6.3-2(b). Typical MTS receiver decoder.

sophisticated low pass filters in the audio circuits that limit the frequency response of each audio channel to 15 kHz (more typically 14 kHz).

Stereo Subcarrier

The stereophonic subchannel is an amplitude modulated, double sideband, suppressed carrier subcarrier. The carrier is suppressed to avoid wasting modulation capability. The subcarrier operates at twice the NTSC video horizontal scanning rate of 15.734 kHz or 31,468 kHz. Station video sync is fed to the encoder to maintain synchronism with the video. The subcarrier is modulated with the stereophonic difference signal or left minus right (L − R). Maximum peak deviation of the main carrier (injection) by the stereo subcarrier energy (not the carrier itself which is suppressed) is 50 kHz.

The modulation level is related to the companding system algorithm and is normally quite high during stereo program material due to the compression. However, because of the nature of stereophonic sound (little correlation between the sum and difference channels) a form of interleaving takes place that, in effect, prevents the deviation of the main carrier from becoming simply the sum of the two components that would appear to produce a total of 75 kHz. Instead, the maximum rarely exceeds 50 kHz. This is because a *left-only or right-only* signal originating from the studio would produce only a 50% modulated signal in the mono channel though it produces a high modulation level in the difference channel.

The difference channel of the BTSC system employs the dbx noise reduction circuit (the compression portion of the compandor) which precisely controls the maximum modulation level of the difference signal. The preemphasis varies according to the level and frequency of the sound entering the compressor. While the combined deviation of main (L + R) and stereo (L − R) carriers rarely produce more than 50 kHz deviation of the main channel, most stereo encoders incorporate some form of peak limiting device in these channels to prevent over deviation.

Note that the difference channel is nonlinear by design. That is, the compandor processes the audio signal differently at low levels than at high levels. When conducting tests on the difference and SAP channels, a 75 μsec preemphasis network is substituted for the compressor. Consult the alignment instructions provided by the manufacturer for specific tests and setup requirements.

The selection of the center frequency of the stereophonic subcarrier is based on its relationship with the 15.734 kHz horizontal frequency (H) component of the video in both the transmitter and the receiver. Crosstalk caused by video signal circuits, power supply and RF paths in the transmitter generally occurs at multiples of "H" and therefore is canceled or at least substantially reduced in the receiver without significantly affecting general stereo audio quality.

The first harmonic of "H" (15.734 kHz), present to some extent in many transmitter and receiver audio circuits, is above the hearing range of most viewers and not normally audible. Receiver audio circuits generally employ notch or low pass filters to eliminate all but vestiges of "H" at the audio output. The second harmonic falls at 2H (31,468.6 Hz) which is the frequency of the BTSC stereo subcarrier. By locking the stereo subcarrier frequency to the video sync, the harmonic and subcarrier remain in phase and no beat occurs.

Modulation of the 2H harmonic by 59.94 Hz field rate sync components of the video signal can also occur in either the transmitter or receiver. The sharp rise time of the 60 Hz component causes the 60 Hz sound to have the characteristic buzz that is heard on some receivers. A related *buzz-beat* effect is reduced with high pass filters in the stereo subchannel in the receiver to remove unwanted audio components below about 100 Hz that can beat with the 59.94 Hz buzz. While this has virtually no effect on the stereo audio quality, it substantially reduces the low frequency buzz and hum caused by the field rate modulation.

Placing the stereo subcarrier, with its 15 kHz sidebands, so close to the upper end of the main audio channel invites potential crosstalk problems. Therefore, it is essential that the subcarrier frequency be an integer multiple of "H" as previously described. Placing the stereo subcarrier at higher or odd multiples of "H" is not desirable due to the much higher noise levels encountered in the upper portion of the aural baseband and the potential for problems with phase locked loop audio detector circuits in the receiver. A higher subcarrier frequency would also have made the addition of the SAP channel more difficult to implement as well.

The buildup of noise in the aural baseband, shown in Figure 6.3-3, illustrates why the stereo subcarrier is placed as low as possible in the baseband and indicates the necessity for the audio companding system. In FM modulation systems, noise increases at a rate of 6 dB for each doubling of the bandwidth. As a result, the noise the stereo subcarrier would add to the received dematrixed sound is about 23 dB. This is reduced about 6 dB because the stereo subcarrier deviates the main carrier by 50 kHz compared to the main channel deviation of 25 kHz. The resulting degradation in signal-to-noise is only about 18 dB. Further, modulation components caused by the horizontal and vertical sync signals also add to the noise buildup. Both interference sources are overcome by the 30 dB BTSC noise reduction system.

Using an AM modulated, double sideband, suppressed carrier subcarrier for the stereo channel also provides several other significant advantages over an FM subcarrier. AM subcarrier technology is well known and developed within the FM broadcast industry. Circuit performance has been raised to near perfect levels. Moreover, AM offers lower theoretical distortion levels than would be possible for FM given the limited spectrum available (30 kHz) for transmission of the stereo subcarrier, and provides an opportunity for future enhancements of the aural channel (by quadrature modulation or other compatible modulation schemes). The 15 kHz frequency response of the stereo subchannel also properly complements that of the main channel.

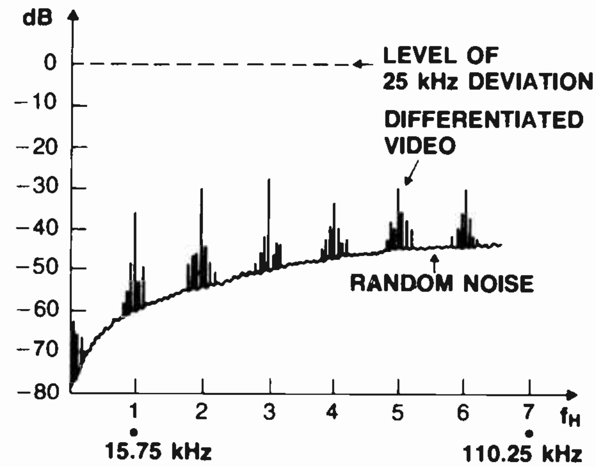


Figure 6.3-3. Spectrum of noise on the received aural baseband. (Courtesy of J.J. Gibson, RCA)

Pilot Carrier

The amplitude modulated stereo subcarrier is transmitted with the carrier suppressed. To provide a reference for the AM detector, an unmodulated pilot signal is transmitted at the TV horizontal line rate of 15.734 kHz that is one-half the subcarrier frequency. The pilot is used in the receiver to reinstate the carrier on the exact frequency and phase as the original. This is similar to the technique employed in broadcast FM stereo transmission systems.

For BTSC stereo, the pilot is locked to the video horizontal sync rate of 15.734 kHz. The pilot modulates the main aural carrier to a deviation of 5 kHz (±0.5 kHz). The stereo encoder requires a feed from the companion video signal to obtain its frequency reference. While it is possible that stereo sound receivers could use the horizontal sync information obtained from the receiver video circuits for the reference, there are receivers that do not process the video when decoding the stereo audio and therefore would not have a reference to synchronize the stereo detector.

The pilot also is used to activate the stereo circuits and indicators in the receiver. Therefore, the pilot must be protected from extraneous signals near the pilot frequency in the aural baseband.

Pilot Carrier Protection

Section 73.682(c)(3) of the FCC Rules requires that modulation components around the BTSC pilot frequency of 15.734 kHz (+ or −20 Hz) be attenuated to 0.125 kHz deviation (46 dB below maximum modulation of 25 kHz). This is the case whether the station is operating with multichannel sound or monophonic sound.

A disadvantage with having the pilot at the horizontal rate is that there may be some residual horizontal sync modulation components and crosstalk in older transmitters or in program audio content that can cause false indications of stereo and improper operation of stereo decoders in BTSC receivers. Crosstalk from horizontal sync can occur in the exciter or amplifiers of older transmitters. Problems of this sort must be solved by broadcasters and manufacturers by adequately shielding or isolating the sync components of the video from the audio circuits in transmitting and receiving equipment. Modifications may be necessary to existing transmission equipment. The level of horizontal sync modulation and crosstalk components in the visual transmitter 4.5 MHz above the visual carrier can be controlled by installing low pass video filters, adjusting transmitter visual RF circuits or improving the 4.5 MHz rejection in the transmitter combiners or diplexers.

Low level sync modulation components with sufficient amplitude to cause false triggering of BTSC receivers can also occur by acoustic pickup by microphones that are in close proximity of video monitors in studios, news sets and in announce booths. Installation of identical high quality low pass or notch filters in the stereo audio path after the last switching point in the broadcast plant will effectively eliminate horizontal sync stray acoustic or magnetic pickup in the studio from interfering with processors, STL facilities and the BTSC encoder at the transmitter. The characteristics of such a low pass filter are shown in Figure 6.3-4. Filters should be installed ahead of devices with preemphasis so as not to compound the problem. For example, a standard 75 μsec preemphasis network has about 15 dB gain at 15 kHz.

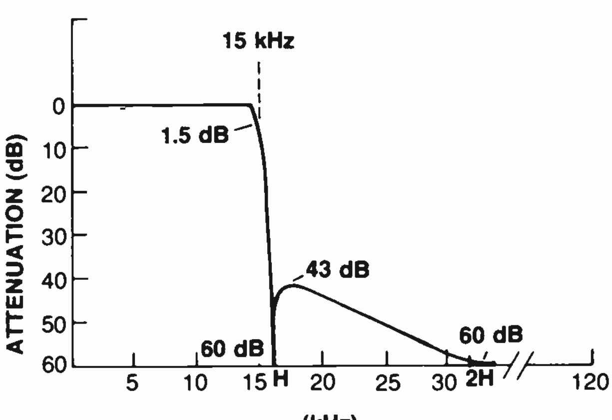


Figure 6.3-4. Audio input low-pass filter.

Operators of monophonic transmitters should also check their facilities for residual horizontal sync signals in either the aural transmitter or in the combining network at the output of the transmitters. The monophonic audio input to the transmitter should incorporate a 15 kHz low pass filter to prevent false triggering of the stereo indicator on television receivers.

The Separate Audio Program Channel

BTSC MTS transmissions may include a separate audio program (SAP) channel in addition to stereo. While the SAP channel may be program related or not, as the station chooses, it is nevertheless designed for reception by the public. If it desired to use the SAP channel for non-public uses, the injection level of the subcarrier must be kept below the receiver threshold (about 5 kHz). If not, viewers with the SAP option switch on their receivers may wonder about the material if it is not related to the video program in some manner.

The SAP channel subcarrier, an FM subcarrier, is located at exactly 5H (78.671 kHz) in the aural baseband (see Figure 6.3-1). Although somewhat noisier at this location in the baseband than if it were placed at 4 H, there is less chance for crosstalk with the main or stereo channels. The same BTSC noise reduction system on this channel provides an adequate signal-to-noise ratio. The audio frequency response is limited to a maximum of 12 kHz (typical response is 10 kHz). The use of the companding system eliminates the need for separate preemphasis in the SAP channel.

The SAP subcarrier is frequency modulated by program material to a maximum deviation of 10 kHz. The subcarrier injection into the main carrier is limited to a peak deviation of 15 kHz. That is, the subcarrier modulates (deviates) the main carrier by 15 kHz. The subcarrier frequency is locked to 5H when not modulated.

The maximum injection level of 15 kHz of the main carrier by the SAP subcarrier is important. Receivers with SAP decoding equipment will detect the presence of the subcarrier and provide either automatic or manual switching and an indicator light. The receiver SAP decoder is designed to activate when the injection level reaches 8 to 10 kHz. Below this level the receiver will ignore the presence of a subcarrier. To use this portion of the baseband for some purpose other than program related audio, a subcarrier with an injection of 3 to 5 kHz would probably be ignored by most SAP receivers. A non-public use of the SAP subcarrier cannot occur at the same time as regular program related SAP channel operation but could be used during stereo or monophonic operation.

The SAP channel uses the same dbx noise reduction scheme as the stereo subcarrier. This permits a single expander decoder to be used in the receiver for either stereo or SAP. Most receivers provide either stereo or SAP output but not both at the same time. To make both available simultaneously a receiver would need two expanders. Alternatively, the SAP output could be provided without expansion but the output would be highly compressed. The sound would be acceptable for some vocal material but would be objectionable for most music.

The Companding System

Companding is a term used frequently in the telecommunication industry to describe what broadcasters call noise reduction, a process well known to studio audio technicians. Its use is restricted to systems where there is control over both ends of the circuit. Examples include magnetic analog audio recorders, microwave and satellite circuits and telephone lines used for program transmission. Simply stated, noise reduction is compressing the dynamic range at the sending end and expanding the signal an equal amount at the receiving end of the circuit. In the process of expansion the noise is reduced. Until the development of the BTSC MTS system, noise reduction techniques could not successfully be used for over-the-air transmission because a complimentary expander is required in the receiver. BTSC MTS required the use of expanders in the receiver from the beginning of the service. As a result, BTSC MTS is the first broadcast service to use companding as an integral part of the transmission system. dbx, Inc. developed the BTSC companding system.

Some FM broadcast stations employ a limited range compression arrangement for an over-the-air noise reduction system. However, because few receivers are equipped with the corresponding expander, the compression must be held to about only 10 dB in order to avoid producing undesirable side effects and becoming objectionable to a listener without a decoder.

Noise reduction was built-in to the BTSC MTS system from the beginning of the service and therefore a high performance system is employed. Figure 6.3-5 graphically depicts how a typical noise reduction system alters the dynamic range of the program material during transmission according to the content. Dynamic

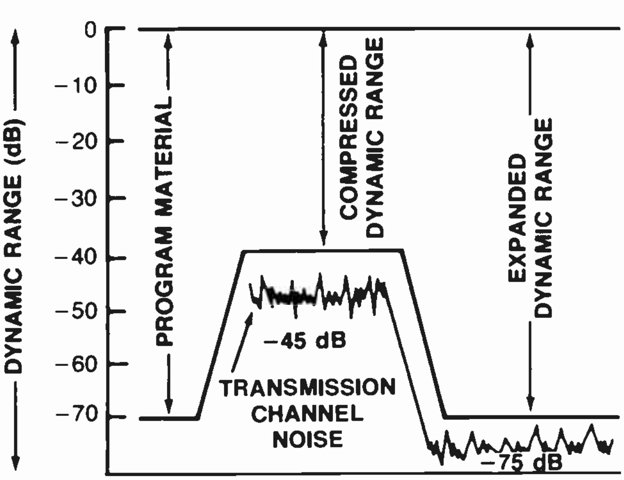


Figure 6.3-5. Typical noise reduction action.

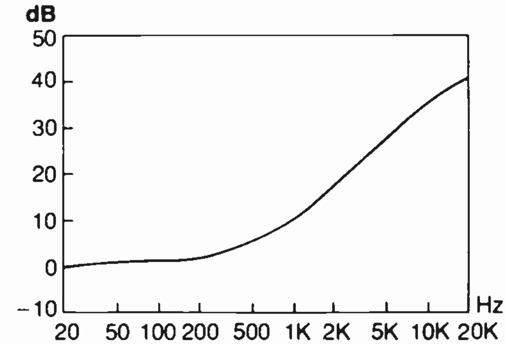


Figure 6.3-6(a). Fixed preemphasis frequency response.

range is reduced from 70 dB to about 40 dB. A matching expander in the receiver restores the dynamic range.

The companding system employed by the BTSC MTS system was specifically designed to operate in the comparatively hostile environments presented by television transmitters, signal propagation, retransmission systems (cable TV, MATV, translators) and the receiver itself. Thermal and impulse noise, multipath distortion and *intercarrier buzz* generated by transmitter and receiver all combine to present formidable obstacles for the companding system to overcome.

The companding system used in the BTSC system employs a combination of fixed preemphasis, *spectral* compression and amplitude compression. The fixed preemphasis combines the familiar 75 μsec rising frequency response with a 390 μsec network. The resulting curve is shown in Figure 6.3-6(a). By itself this extremely steep preemphasis curve would cause problems with high audio frequencies. To avoid this, a dynamic spectral compressor (variable preemphasis) circuit is employed to increase the gain of low level, high frequency material and reduce the gain of high level, high frequency audio. Only frequencies above about 200 Hz are affected. High frequency material is increased during transmission by as much as 30 dB at low levels but virtually not at all at high levels. Figure 6.3-6(b) illustrates the dynamic characteristics of the spectral compressor. A complicated algorithm controls

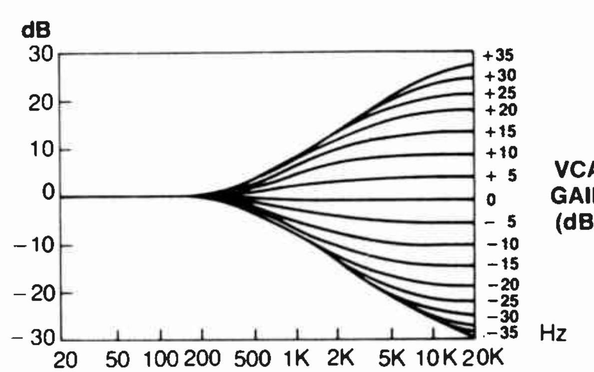


Figure 6.3-6(b). Spectral compressor, frequency-response range.

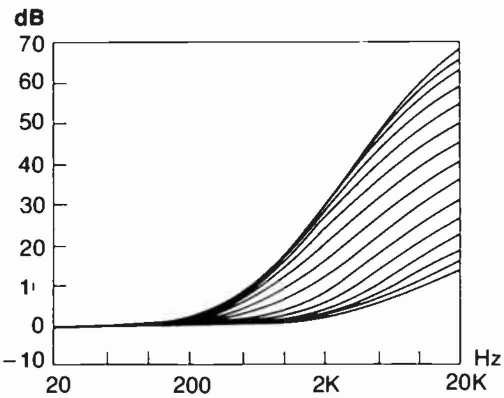


Figure 6.3-6(c). Frequency response range of spectral compressor and fixed preemphasis.

the fixed and variable preemphasis to produce the sharply rising frequency response shown in Figure 6.3-6(c). The third feature of the companding system is the amplitude compressor which reduces the dynamic range of the input signal by a factor of 2:1 for low frequencies, and 3:1 for high frequencies. In other words, a 40 dB dynamic range audio signal is reduced to 15 to 20 dB or less for transmission as shown in Figure 6.3-7. In addition to dynamic range compression, the maximum input level applied to the stereo subcarrier is reduced to just below the maximum modulation of the subcarrier. This feature provides some headroom to accommodate instantaneous peaks without resorting to clipping. High level transients are clipped to avoid severe distortion during reconstruction of the compressed signal in the receiver expander.

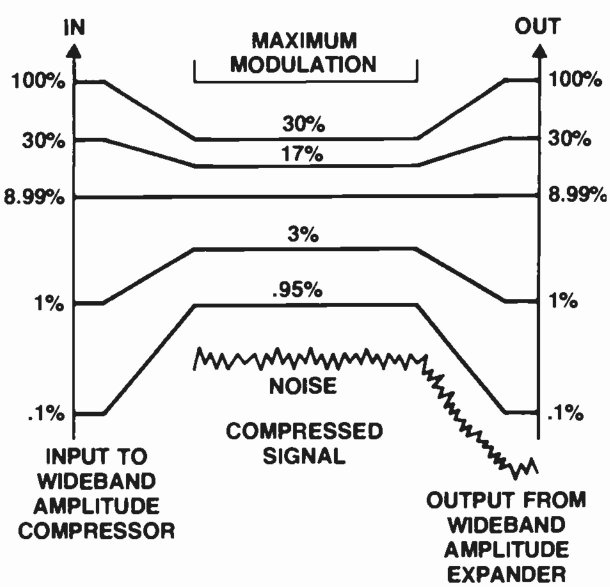


Figure 6.3-7. Companding of (L − R) and SAP. (Courtesy of dbx, Inc. dbx is a registered trademark of Harmon International, Sandy, Utah, USA.)

The equivalent 75 μsec mode in the companding system uses a precision 75 μsec network in place of the nonlinear dbx compressor when tests for separation, crosstalk and noise tests conducted. In this mode the exciter contains only linear elements which allows routine adjustments on the exciter and transmitter and measurements of the main, stereo and the SAP channels to be performed more accurately and very much as conventional FM broadcast stereo exciters. CEMA publishes the *Compandor Complexity Analyses* that supplements *OET Bulletin #5* described earlier.

The use of compandors and noise reduction systems has gained wide acceptance throughout the broadcast industry. It is not unusual for any given program to undergo noise reduction processing several times by the time it reaches the viewer. The mistracking of a single companding system in the chain can result in substantial errors in the resulting sound that can be more objectionable than the problems the compandor was designed to solve. In order to prevent audible artifacts from occurring, only the highest quality processing and noise reduction systems should be employed and these must be correctly operated and properly maintained.

Professional Use Channel

The PRO channel, which is not designed to be received by the public, can be used by broadcasters for many purposes such as voice cues, signaling and data transmission. The PRO channel is located at 6.5H (102.28 kHz, ±500 Hz) in the BTSC baseband as shown in Figure 6.3-1. The carrier need not be locked to "H". The peak deviation of the frequency modulated subcarrier is 3 kHz. The peak deviation (injection) of the main carrier by the PRO subcarrier is 3 kHz. When used for voice or audio transmission, a modulation limiter, 3.4 kHz bandlimiting filter and a 150 μsec preemphasis network are normally employed. For data transmission the limiter, preemphasis network and filter are bypassed.

Other Subcarriers

In addition to providing for stereo and SAP subcarriers, the FCC permits television stations to use any other scheme of subcarriers on the aural carrier as long as they do not interfere with the normal operation of monophonic or BTSC television receivers. When not transmitting stereo or SAP, broadcasters may elect to use the aural baseband for subcarriers for non-program related purposes. Virtually any kind of subcarrier for program material or data may be transmitted using AM, FM, other analog or digital modulation methods. Subcarriers may be placed in the aural baseband between 16 and 120 kHz but must avoid the BTSC pilot and SAP frequencies. While any number of subcarriers may be used, the maximum deviation of the main carrier by these subcarriers is limited to a total of 50 kHz. The total modulation of the aural carrier including main channel sound and non-BTSC subcarriers may not exceed 75 kHz.

Stations should avoid placing non-public subcarriers

Table 6.3-2
Typical modulation monitor functions.

Signal	Monitored Parameter
Main channel (L + R)	25 kHz peak deviation
Stereo channel (L + R)	50 kHz peak deviation
Main and stereo	50 kHz peak deviation
Pilot carrier injection	5 kHz peak deviation
SAP modulation	10 kHz peak deviation
SAP injection	15 kHz peak deviation
Other subcarrier	3 kHz peak deviation
Total composite	73 kHz peak deviation

within ±10 kHz of the SAP subcarrier frequency unless the injection level of the subcarrier is kept well below 5 kHz deviation. The use of a single subcarrier using the full 50 kHz deviation should be avoided because of the potential of crosstalk into the main channel in some receivers unless extensive compatibility testing is conducted. Multipath in the propagation path can exacerbate crosstalk with large subcarriers.

Modulation Summary

The FCC permits the BTSC MTS signal to modulate the aural carrier a total peak deviation of 73 kHz. This is produced by the combined main monophonic and stereo subcarrier deviation of 50 kHz, the pilot subcarrier deviation of 5 kHz, the SAP subcarrier deviation of 15 kHz and the PRO subcarrier of 3 kHz for a total of 73 kHz (see OET-60A section D(b)(6)). Table 6.3-2 lists the MTS modulation components that should be monitored regularly.

Other subcarrier arrangements may modulate the main carrier to an additional peak deviation of 50 kHz. The combined main monaural channel of 25 kHz and non-BTSC subcarriers of 50 kHz shall not exceed 75 kHz deviation (see Section 73.68(c)(9) of the FCC CFR).

TRANSMISSION REQUIREMENTS

Many of the technical rules in Part 73 of the FCC CFR that described quality characteristics of the transmitted signal were deleted by the FCC in 1984. Those rules, requiring stations to meet certain distortion, frequency response and signal-to-noise performance levels, were eliminated in favor of marketplace pressures on stations to maintain high quality levels. Thus, there are no FCC rules which require minimum audio quality performance levels in the AM, FM or TV broadcast service. The FCC Rules that permit stations to transmit MTS also do not specify performance objectives. However, rules which serve to control interference have been maintained.

In addition to FCC and CEMA publications transmitter and BTSC encoder and modulation monitor manufacturers also publish booklets and manuals describing potential problems encountered in preparing transmission systems of BTSC audio along with methods for measurement and problem correction. Some of these are listed in the references at the end of this chapter.

The BTSC MTS performance objectives achievable under normal operating conditions are shown in Table 6.3-1. While most of the objectives indicate the potential for high audio quality, the stereo separation figure may be of some concern to those familiar with the much higher performance of the FM stereo system. While more than 30 dB separation can be achieved in practice with BTSC, separation values greater than 30 dB produce little subjective improvement. The nonlinear circuits in the companding system and the critical adjustments which must be made in the BTSC encoding system may result in less than the typical 40 dB or more separation normally achieved in broadcast FM stereo operation. The tracking gain between the mono and stereo channels is the single most important operating characteristic that determines the relative separation between channels as shown in Figure 6.3-8a.

While it may be possible to balance the gain to a tenth of a dB between the sum and difference channels in FM radio stereo transmission, a balance is about 0.5 to 0.8 dB is typical for the BTSC system. This is because the noise reduction system compresses the stereo channel by as much as 30 dB, which magnifies the difference when received, decoded and expanded in the TV receiver. However, with high stability circuits and attention to balance 40 dB stereo separation can be obtained.

Phase shift between audio channels also reduces stereo separation. Only a few degrees difference can reduce separation to 30 dB or less as shown in as shown in Figure 6.3-8b. When combined with channel gain differences, separation will be further reduced. Incorrect stereo channel carrier phase error and stereo pilot phase error also contribute to reduced stereo separation as shown in Figure 6.3-8c.

A major feature of the BTSC MTS system is that there is no stereo noise penalty as there is with FM stereo radio. Because the FM stereo radio subchannel has no companding it is subject to the increased noise in the channel baseband by as much as 23 dB. In FM radio the main or L + R channel must also be reduced by 8 to 10% to accommodate the addition of the pilot subcarrier. With BTSC MTS, the aural carrier deviation is increased when adding the pilot and stereo subcarriers. Thus, there is no reduction in deviation of the mono signal during BTSC stereo transmission as there is when adding stereo in broadcast FM radio operation.

The crosstalk performance values are the single most important technical parameters with which the broadcaster must be concerned for the BTSC system. Most crosstalk is caused by components and systems external to the stereo transmission equipment itself. Nonlinearities in the TV aural exciter and power amplifier stages, combining networks, antenna switching components and the antenna itself are locations in the signal path where crosstalk can be generated. In general, crosstalk into the stereo signals from the SAP will be less than crosstalk from the main channel into the SAP signal. As long as crosstalk levels do not exceed the

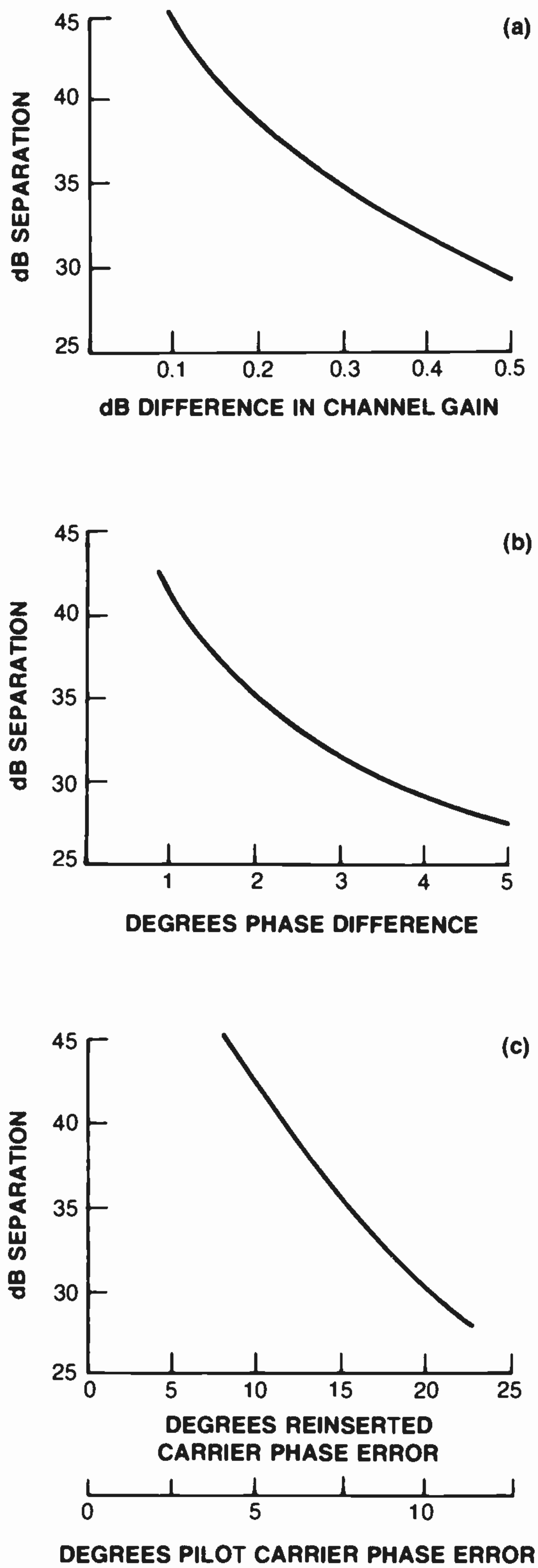


Figure 6.3-8. Stereo separation versus (a) amplitude difference between main and stereo channels, (b) phase difference between main and stereo channels, and (c) error in phase of pilot or reinserted carrier.

CROSSTALK	OBJECTIVES
MAIN INTO STEREO	>40 dB
STEREO INTO MAIN	>40 dB
MAIN INTO SAP	>50 dB
SAP INTO MAIN	>60 dB
OTHER SUB INTO MAIN	>60 dB
OTHER SUB INTO STEREO	>60 dB

Figure 6.3-9. Crosstalk objectives.

levels shown in Figure 6.3-9, most crosstalk will be concealed by the companding system.

Other sources of crosstalk are multipath distortion due to poor transmission line and antenna VSWR, multipath (ghosts) during the propagation of the transmitted wave and improper matching in the receiving antenna system. In general, the companding system can conceal noise and crosstalk in the stereo and SAP channels subjectively by as much as 30 dB.

Setting Modulation Levels

The most important means for maintaining good stereo separation is to correctly set modulation levels of the L − R stereo subcarrier. Errors of less than 0.5 dB will result in a reduction of stereo separation to 30 dB, the FCC minimum. For best performance, levels must be set to an accuracy of 0.3 dB to help account for crosstalk problems in other parts of the system. While levels can be estimated with modulation monitors, the best method for setting levels to this accuracy is to use the Bessel null procedure with a highly accurate, low distortion audio test signal and an RF spectrum analyzer.

Visual Over-Modulation

One way to reduce buzz in television receivers is for the broadcaster to insure that the video modulation level is never allowed to exceed the FCC maximum of 87.5%. Video levels that vary widely during program production and transients from character and graphics generators can easily cause the visual modulation level to approach 100% or carrier cut-off. At this point the aural detector in intercarrier type television receiver loses its reference (the aural carrier) and the result is a strong 59.94 Hz buzz in the received audio. Proper adjustment of video processing equipment will eliminate this problem as a source of buzz. In some cases it may be necessary to add high quality low pass filters to the video outputs of character and graphics generators to limit the high frequency content.

Incidental Carrier Phase Modulation

The control of ICPM in the visual transmitter is also an important performance characteristic of the transmitter that will affect the quality of multichannel sound.

ICPM in visual transmitters must be adjusted and maintained to less than 3° during the luminance portion of the video and less than 5° during sync in order to keep buzz to acceptable levels as shown in Table 6.3-3.

Table 6.3-3
Intercarrier buzz and buzzbeat levels.
(Courtesy of J.J. Gibson, RCA)

Degrees ICPM →	1°	2°	3°	4°	5°
Baseband buzz level in dB below 25 kHz dev. ("worst case")	−50	−44	−40	−38	−36
Stereo SBR dB	56	50	46	44	42
SAP buzzbeat THD dB	−49	−43	−39	−37	−35

Nyquist slope eqv. ICPM = $\begin{cases} 2.4° \text{ at (L − R) subc.} \\ 6° \text{ at SAP subc.} \end{cases}$

Excessive ICPM in the transmitter will cause an audible and annoying buzz in intercarrier sound television receivers. The buzz will be modulated by the action of the expander and increase the level of annoyance.

Intercarrier type receivers, typical of most receivers, are used for multichannel sound reception because of their high immunity to common mode phase modulation that can occur in non-linear active devices or distribution and conversion systems that are found in the path of the television signal before it reaches the receiver. Common mode phase modulation often occurs in TV translators, MATV downconverters, and cable television headend and set-top converter equipment.

In intercarrier type receivers, the aural and visual carriers beat together in the video detector to produce the 4.5 MHz aural IF signal. Because common mode phase modulation affects both carriers, the phase modulation is factored out of the 4.5 MHz carrier. Therefore, ICPM that occurs on the visual carrier in the transmitter, which is independent of the aural carrier, will produce undesired phase modulation of the 4.5 MHz aural carrier in the receiver and the result will be the familiar buzz.

Phase modulation in the local oscillator of up and down converters can be caused by the power supply or mechanical vibration or if the phase of the composite signal is changed by a non-linear amplifier. Phase modulation is common mode when both aural and visual carriers are affected equally. This would be the case when the aural signal is carried as a subcarrier on the visual carrier rather than treated as separate signals as is the case in many broadcast transmitters.

Intercarrier receivers carry the aural signal to the detector as a subcarrier of the visual carrier. Translators and most cable television systems carry the aural through headend processors as a subcarrier. Therefore, if phase modulation is introduced in these systems, it is common mode and will be rejected by the receiver. Intercarrier sound receivers cannot reject phase modulation that is introduced independently into the visual or aural carriers at the transmitter or in subsequent equipment or systems.

Phase modulation in the visual transmitter can be caused by hum in the carrier frequency or upconverter oscillators and by non-linearities in the visual modulator, IF and power amplifier stages. Phase correction circuits in the video modulator can compensate for most ICPM problems. UHF transmitters which use klystron anode pulsers to achieve higher operating efficiencies may have phase modulation introduced by the pulser modulating the visual amplifier power supply at the video field rate (59.94 Hz) and the line rate (15.734 kHz). The high frequency component is not heard by most viewers but the low frequency component is heard by all as a raspy buzz. An anode pulser on the visual amplifier can cause crosstalk through the common power supply into the aural amplifier. The resulting 15.734 kHz modulation, if high enough, can cause false indications of the stereo indicator on BTSC MTS television receivers.

Measuring and Correcting ICPM

ICPM is manifested as differential phase in the demodulated video. Measurement instrumentation requires a TV demodulator with both envelope and quadrature video detectors and a waveform monitor with the ICPM graticules. After measuring ICPM or listening for buzz on a stereo television receiver to determine if ICPM is present, phase adjustments in the modulator or upconverter and retuning the RF stages will result in substantial improvements. ICPM adjustments, if provided on the transmitter exciter, will also affect video differential phase. The ICPM problem should be addressed and solved first, then the video correction circuits can be adjusted to compensate for differential phase.

ICPM is more noticeable on television receivers equipped for BTSC stereo sound because ICPM affects subcarriers to a greater extent than the main channel. In addition, improved audio performance in the new multichannel receivers makes existing buzz more noticeable. Therefore, stations that implement BTSC multichannel sound should check their transmitters for ICPM with an ear toward improving the quality of the mono sound on mono receivers as well as stereo on new MTS receivers.

Synchronous Amplitude Modulation

Synchronous amplitude modulation (SAM), occurs when the frequency swing of the aural carrier is greater than the flattest portion of the tuned circuits through which the carrier passes. A reduction of the carrier occurs where the passband begins to roll-off as shown in Figure 6.3-10. The effect is to cause amplitude changes to the RF carrier during FM modulation. The SAM is in sync with the FM, hence the name synchronous AM. SAM is made worse if the aural RF carrier is not centered within the passband of the tuned circuits in the aural transmitters, combiners and diplexers. While most TV transmitter systems provide adequate aural RF bandwidth, the position of the carrier within the passband is critical and should be checked regularly.

SAM manifests itself as intermodulation distortion which in turn causes an increase in crosstalk between

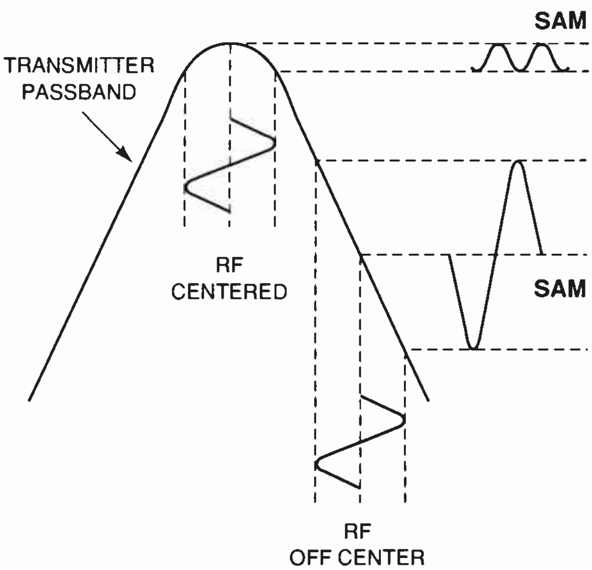


Figure 6.3-10. Synchronous Amplitude Modulation (SAM).

channels and decreased stereo separation. The SAP and PRO channels are more susceptible to SAM than the main channel especially during high levels of main channel modulation when the frequency swing is greatest.

Compatibility with Monophonic Receivers

During the development of the BTSC MTS considerable effort was spent evaluating the effect of a fully loaded BTSC system on existing monophonic television receivers. Extensive compatibility testing of a wide range of receiver makes and models, under adverse transmission conditions, revealed that the BTSC System caused virtually no significant degradation to either the sound or picture of the desired or adjacent channel.

BTSC sound transmission will not introduce additional buzz or distortion in existing monophonic receivers. On the other hand, non-linear circuits and the resulting amplitude modulation to phase modulation (AM-to-PM) conversion will nearly always produce buzz which will be more objectional in stereo than in mono. Most of the buzz heard in mono or stereo receivers is usually created in the transmitter or the cable TV system in the form of ICPM or video over-modulation.

Modern monophonic television receivers should not experience significant degradation when receiving a BTSC MTS signal. In fact, monophonic receivers may actually exhibit a noticeable improvement in performance when tuned to BTSC stereo transmissions due to the improved performance of the stereo exciter, broadband aural RF systems and lower ICPM in visual transmitter components.

These desirable side benefits easily offset the slight reduction in monophonic modulation level which occurs when transmitting a mostly left or right channel sound. Because the main channel is the sum of left plus right, a right-only or left-only signal will produce up to 6 dB less modulation than the sum signal. However, program producers seldom employ left or right-only signals except for sound effects and music. Virtually all dialog is placed in the center (L + R) so there will be no reduction in modulation.

Compatibility with Cable Television Systems

Most broadcast television signals are carried by cable television systems. In 1997, about 65% of TV households had at least one receiver connected to a cable system. It is important, therefore, that broadcasters be concerned with the quality of the audio as well as the video signal as carried on cable television systems.

The BTSC MTS system was specifically designed to pass through properly adjusted cable television and master antenna TV (MATV) headend and distribution equipment and subscriber terminals. The concern that the lower sideband of the upper adjacent visual carrier might affect the wider BTSC MTS modulated aural carrier of the lower channel has not been confirmed in practice. In general, those cable TV systems which process off-air signals in an RF mode (that is, where the aural carrier is not demodulated) will have little difficulty in successfully passing an acceptable BTSC MTS signal through the system. The same is true of the RF conversion type set-top converters used by CATV systems.

On the other hand, cable systems with demodulating type headend or subscriber terminal equipment may not be able to pass MTS signals without substantial degradation to the stereo signal due to the use of low pass filters in the composite audio portions of demodulator facilities where stereo and SAP subcarriers are filtered out or severely distorted. However, newer cable headend equipment process BTSC stereo with very little degradation by carrying the BTSC composite signal through at IF (such as 4.5 MHz).

The baseband type set-top converter poses the greatest threat to BTSC stereo if the audio is demodulated within the converter itself. This is often done to permit the audio level, as well as the channel, to be controlled by a remote control device. BTSC stereo will not reliably pass through such devices and both stereo and SAP signals will be lost.

Most cable television systems maintain the visual-to-aural carrier ratio at 15 to 17 dB compared with the normal 7 to 10 dB for broadcast. This can add 5 to 10 dB more noise and buzz to the BTSC signal in the cable system. Therefore, it is important that the broadcaster limit degradation to the broadcast BTSC signal in order to insure that stereo carriage on cable systems is satisfactory.

The BTSC MTS signal should pass, without significant degradation, through amplitude and frequency modulated link (AML, FML) microwave systems or fiber optic systems used by many cable companies for wide area distribution of multiple channels.

Cable companies also routinely convert discrete stereo audio received from satellite program sources into BTSC stereo in NTSC modulators for distribution to subscribers. In some instances, less than full BTSC performance is achieved from these typically low cost modulators.

Television broadcasters will find it advantageous to determine how their stereo sound is processed in cable TV systems on which their station is carried and conduct routine performance tests. Most television stations routinely check the quality of the visual signal on cable TV systems which can introduce several additional active elements between the broadcast transmitter and home receiver. Basic audio tests for noise, stereo separation, pilot level, SAP channel quality and crosstalk between SAP and main channels should, at the very least, be conducted semi-annually at the output of a subscriber terminal on the largest cable systems carrying the station. This procedure will help insure that the cable company is properly handling (quality and loudness) the broadcast signal as well as cable satellite channels. A BTSC receiver of known characteristics and standard audio test instrumentation is needed to conduct the tests on the BTSC MTS signal. Such tests can be conducted with ordinary signal generators and audio level meters.

Compatibility with Television Translators

Many television broadcast signals are carried to remote communities by the use of translators that use RF power amplifiers ranging in size from 1 W to 1 kW.

Television translators that use heterodyne or direct RF conversion techniques normally do not cause degradation to BTSC MTS signals if the translator is properly maintained. Some translators, however, receive little attention unless they fail completely or viewers report poor performance.

Translators that separately process the visual and aural carriers may require modifications or adjustments to accommodate the BTSC MTS signal. Improvements include a wider bandwidth for the aural signal to accept the subcarriers in the BTSC signal, good visual transmitter oscillator stability to reduce ICPM and correct RF tuning to reduce SAM. Remodulating translators that demodulate and process the video also process the audio as a 4.5 MHz subcarrier similar to cable TV systems and have improved RF bandwidth characteristics with the result that little degradation occurs to the BTSC signal. Reducing the BTSC signal to wideband audio baseband with subcarriers requires a very linear FM modulator in order to accommodate the complex BTSC baseband.

Modulation Monitoring

Monitoring the modulation and conducting routine tests of the BTSC system can be accomplished by either a spectrum analyzer plus the standard audio instrumentation or a BTSC modulation monitor which provides continuous modulation information on all critical operating parameters. Stations must have the ability to monitor main mono channel, stereo composite, pilot, SAP, (and PRO channel, if used) and overall aural modulation levels during normal operation. Virtually all commercial modulation monitors provide the means to make accurate measurements of all operating parameters and technical tests. Table 6.3-2 lists the regularly monitored MTS transmission parameters.

Operating Practices

A compilation of transmission impairments that can cause stereo audio impairments is shown in Table 6.3-4.

Stereo synthesizers, used by some stations to provide a sense of stereo during monophonic programming, should not be used in place of or in addition to stereo encoding. Synthesizers can cause loss of localization of stereo sound sources because some program material accurately generates center channel information. Further, the effects of surround sound and other special aural techniques may be lost if the synthesizer is employed.

Table 6.3-4
Effects of transmission impairments.
(Courtesy of J.J. Gibson, RCA)

Effect / Cause	Buzz	SAP Buzz Beat	Noise & Noise Pumping	Reduced Stereo SEP	Dist.	X-Talk
Non-linear PIX	✓					
Non-linear sound					✓	✓
Mismatches						✓
Diplex filter					✓	✓
ICPM	✓	✓				
Multipath	✓	✓	✓	✓	✓	✓
Weak signals			✓	✓		
Co-channel					✓	✓
Mistracking				✓		
Rec. filters	✓	✓		✓	✓	✓
FM detect	✓				✓	✓

ICPM, visual carrier over-modulation, severe group delay and phase distortion in the visual amplifier will cause buzz in the received audio and loss of stereo separation. These two characteristics are the most sensitive to problems with the stereo generator and RF systems.

Adjustment of the BTSC stereo encoder and dbx compressor circuits (used in both stereo and SAP channels) is not recommended without a thorough understanding of the overall system and appropriate test equipment. The service manuals for the stereo exciter contain the specific procedures for routine maintenance.

Ordinary harmonic distortion in the BTSC stereo baseband (50 Hz to 120 kHz) will result in increased crosstalk between channels. In order to keep crosstalk within limits distortion must be maintained to less than 0.5%.

FCC rules require all stations to provide protection to the pilot frequency by limiting modulation in the vicinity of the pilot 15,734 Hz (±20 Hz) to no more than 46 dB below 50 kHz deviation. The BTSC pilot frequency is immediately adjacent to the upper end of the frequency response for the main L+R channel. It is essential that monophonic as well as stereo stations insure the BTSC pilot is protected to avoid causing improper operation of stereo receivers and false stereo indications. While all MTS stereo exciters incorporate low pass filters with a sharp cut-off above 14 kHz, monophonic stations should install such a filter if adequate protection is not otherwise provided (see Figure 6.3-4). Measuring the pilot protection can be performed with a BTSC modulation monitor or a spectrum analyzer.

Many stations prefer to locate the stereo encoding equipment at the studio and transmit a composite MTS signal to the transmitter over an STL. The STL in effect becomes another section of the RF path where degradation to the BTSC signal can occur. Stereo separation, a good indicator of overall performance, will suffer if the baseband BTSC signal is degraded. In order to maintain at least 30 dB separation at the output of the transmitter, the STL should have at least 40 dB separation to compensate for degradation in the transmitter and elsewhere in the RF path. This means that the phase and amplitude frequency response across the 73 kHz baseband of the STL must be maintained to less than 0.1 dB combined with no more than 1° phase error as shown in Figure 6.3-11. This level of performance is achievable with modern BTSC encoding equipment.

For good stereo operation it is necessary that the left and right channels remain well separated ahead of, within, and at the end of the transmission chain.

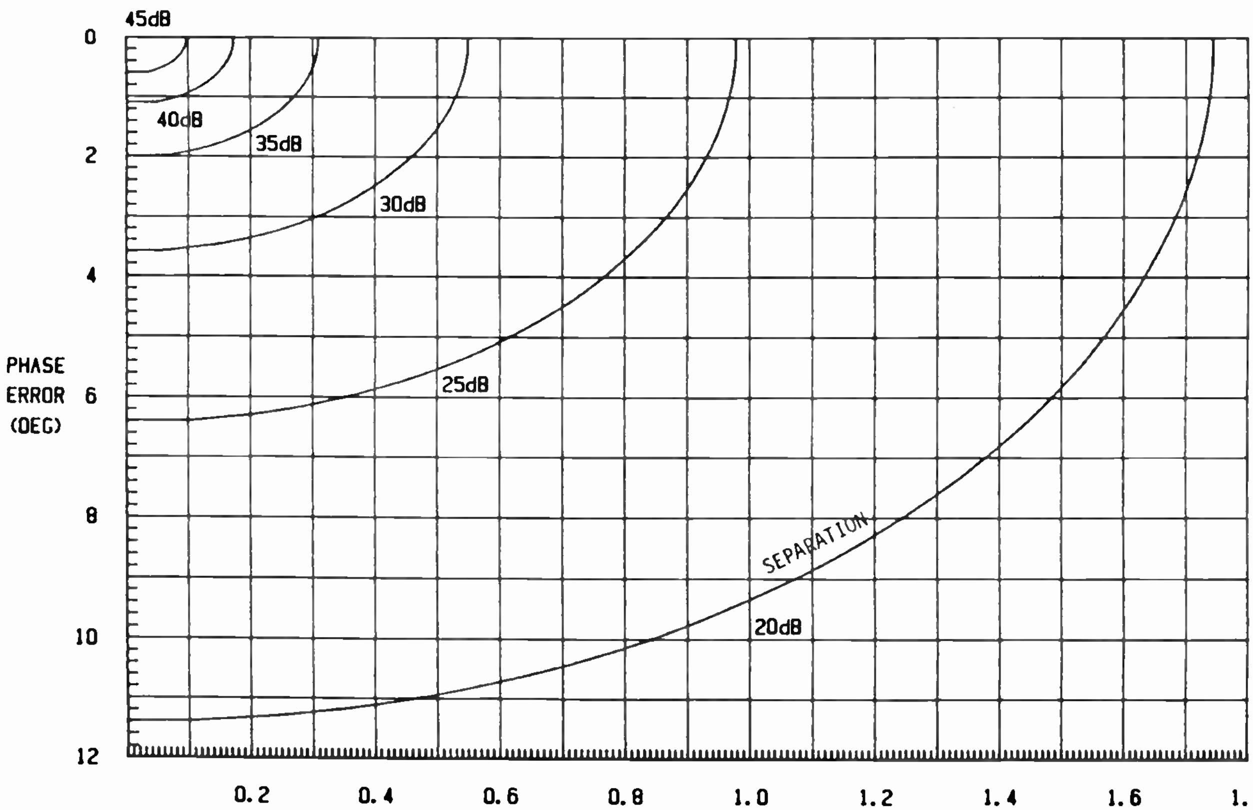


Figure 6.3-11. BTSC Separation vs. combined amplitude and phase errors in the composite baseband. (Courtesy Broadcast Electronics, Inc.)

The amplitude and phase of the left and right signals must be nearly identical and the phase of the pilot carrier must be maintained. Channel separation as a function of these three factors is given in the equation in Figure 6.3-12:

Separation in dB

$$= 20 \log 10 \frac{[(\cos \varphi + S/M \cos \Phi)2 + (\sin \varphi)2]}{[(\cos \varphi - S/M \cos \Phi)2 + (\sin \varphi)2]}$$

Where:

- M = the gain of the main (L + R) path
- S = the gain of the stereo (L − R) path
- Φ = the phase error of the reinserted 2H subcarrier (twice the error of the 1H pilot carrier)
- φ = the difference in phase between the L + R and L − R paths

Figure 6.3-12. Formula for calculating phase and amplitude changes.

Total Harmonic Distortion (THD) and intermodulation distortion on the composite STL must not exceed 0.25% to maintain crosstalk between channels at acceptable levels. If the SAP and PRO channels are also included in the composite baseband, the baseband performance values cited above must extend to 106 kHz.

Placing discrete stereo audio channels on the STL remains popular because most stereo encoders are located at the transmitter. Variations in the performance of discrete audio channels will produce less degradation of the stereo signal than will the same performance STL characteristics on the BTSC composite signal.

The BTSC stereo system has not become attractive for use on satellite transmission circuits for the reasons cited above for composite STL systems.

Off-air monitoring is essential to the maintenance of good stereo and monophonic sound. Reversed channel polarity (sometimes called *phase reversal*) on one channel may cause the stereo sound to be slightly degraded on stereo receivers but severe degradation will result to the sound on monophonic receivers. By using a consumer receiver for audio monitoring, buzz caused by over-modulation or ICPM can be detected at the station. Many stations employ both stereo and monophonic monitoring of the transmitted audio in order to insure that stereo imaging and channel polarity are correct.

The implementation of multichannel sound in a television facility which has not dealt with stereo or multiple audio channels in the past will find the experience to be a challenge. Virtually all audio equipment in the station will need to be reviewed, reworked, rewired, or replaced when adding new audio channels. Videotape machines, routing switchers, master control switching and monitoring, studio audio facilities and the STL will require close examination to determine how to implement extra sound channels. Videotape equipment with two sound channels may require new test procedures to determine that the phase relationships between channels is correct.

Special attention must be paid to audio processing. There will be a tendency towards maximizing the loudness of the signal rather than moderate processing which will result in more acceptable stereo sound quality. Sounding louder than the competition is fashionable in FM radio broadcasting but may not be necessary for television.

Multichannel sound equipped television receivers have substantially improved audio systems which will reveal audio faults now concealed by most television receivers. As a result, special attention must be devoted to the audio transmission facilities of all television stations, whether transmitting monophonic or multichannel sound.

REFERENCES

Multichannel Television Sound Transmission and Audio Processing for the BTSC System, OET-60A, Federal Communications Commission, 1986.

Multichannel Television Sound BTSC System Recommended Practices, Television Bulletin #5, Consumer Electronic Manufacturers Association, 1985.

Small, Eric, *Understanding MTS Equivalent Mode, and Is Your Transmitter Stereo-Ready?* Modulation Sciences, Inc.

Mendenhall, Geoffrey N., *Testing Television Transmission Systems For Multichannel Sound Compatibility,* Broadcast Electronics, Inc.

Carpenter, James R., *Demystifying TV MTS Specifications,* Broadcast Electronics, Inc.

BTSC Stereo: TV Aural Proof of Performance Guide, TFT, Inc.

Evaluating BTSC Stereo Generators and Audio Processors: Some Suggestions by Robert Orban, Orban Associates, Inc.

6.4
TELEVISION DATA BROADCASTING

KELLY T. WILLIAMS
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC
NEIL MITCHELL
PHILIPS SEMICONDUCTORS, SUNNYVALE, CA

INTRODUCTION

Changes in telecommunications markets and technologies are challenging TV broadcasters to broaden the definition of their business and address new market opportunities. One of these opportunities, which has excellent market potential, is the business of *data broadcasting* or *datacasting*.

The datacasting business is on the brink of becoming an explosive growth market. This is evidenced by the growth in demand for traditionally wireline based services such as the Internet and on-line services. With the advent of mobile computing, users will continue to want to have access to information services, but without connecting to a telephone line. Even in a home or business environment there are advantages to a wireless connection to information services, such as no requirement for additional phone lines and ease of installation. Datacasting promises to be a useful opportunity for television broadcasters looking to respond to the demand for data based information services.

The purpose of this chapter is to provide a fundamental understanding of the regulatory and technical issues of data broadcasting and to look at some of the potential applications for digital television. It begins with an overview of datacasting and its general applications, reviews regulatory issues and surveys the National Television Systems Committee (NTSC) technologies that are currently authorized by the FCC. The second section of this chapter examines specific datacasting applications that could be supported by the Advanced Television Systems Committee (ATSC) digital television system.

NTSC DATACASTING

What is Datacasting?

Generally, datacasting can be defined as the simultaneous delivery of digital information from a central source to a large number of geographically dispersed receivers. Specifically, current television datacasting uses the NTSC signal as the delivery mechanism for data based information. One very basic example of this is the closed caption system which sends digital information via the station's vertical blanking interval (VBI), at a modest rate of 480 bits per second (bps), to special decoders. Once decoded, the information is displayed on the TV screen as text.

Datacasting Applications

Television data broadcasting can be used to provide a multitude of services to businesses and consumers. Some of those applications are:

- Station information (call letters, network affiliation, etc.)
- Station program information (program schedules)
- Sports information (statistics)
- Traffic and weather information
- Internet information (news group and e-mail delivery)
- Electronic billboard content information
- Public utility load management

These are examples of applications that broadcasters can initiate themselves. There are additional applications and services that broadcasters could provide in partnership with other content providers, such as:

- Financial information (stock services)
- News services (newspapers, directed industry news)
- Web pages
- Interactive software
- Specialized business services (credit card verification data)

The potential for developing applications for data broadcasting is almost endless. So long as there is a demand for some type of digital information, TV data broadcasting can provide an innovative solution for wireless delivery.

Datacasting Regulations

The Federal Communications Commission (FCC) has allowed VBI based datacasting for some time. In addition to closed captions (Section 73.682(a)(21) of the FCC's Rules) there has been a general provision in the Rules for the *Telecommunications Service on the Vertical Blanking Interval* since March 4, 1985. As such, it is now general knowledge that broadcasters can place information, either analog or digital in nature, in their VBI without prior authorization from the FCC. This provision is found in Section 73.646 of the FCC Rules.

On June 21, 1996, the FCC adopted a *Report and Order* in MM Docket No. 95-42 that expanded TV broadcast stations' authority to use certain datacasting technologies without seeking the prior approval from the Commission. The FCC amended its rules to allow TV stations to transmit ancillary data in the NTSC television signal. The Commission approved systems from four companies for use by TV broadcasters: Yes! Entertainment Corporation, A. C. Nielsen Company, Digideck, Inc. and WavePhore Inc.

The FCC started this proceeding in May 1995, seeking to determine how best to permit data broadcasting technologies to be integrated into the NTSC broadcast television service. The *Notice of Proposed Rule Making* (NPRM) was initiated partially in response to a letter to the Commission from WavePhore requesting a declaratory ruling that licensees may, without prior Commission authorization, use WavePhore technology to transmit digital data signals.

The Commission, however, also had a number of unresolved requests for authorization that it bundled into this proceeding. In addition to the four companies listed previously, this proceeding also considered requests from Airtrax, Incorporated and En Technology. However, the Commission only approved the previously listed companies in the *Report and Order*. The FCC did subsequently approve the use of En Technology's system but only for limited use. The FCC also later approved a datacasting system from Microsoft.

FCC Approved Technologies for NTSC

A.C. Nielsen's AMOL System

Automated Measurement of Lineup (AMOL) is a proprietary VBI based technology that operates on line 22, the first line of active video. In the *Report and Order* in MM Docket No. 95-42, the FCC granted permanent authority to Nielsen to operate its system on line 22, but on a non-exclusive basis. Previously, Nielsen had been operating on a temporary, conditional authority issued by the FCC in November 1989.

Digideck, Incorporated

Digideck was a participant in the National Data Broadcasting Committee (NDBC) evaluation program. It was through this participation that the Commission approved its technology for use by broadcasters. The NPRM in MM Docket No. 95-42 did not discuss Digideck's technology; however, the work of the NDBC was submitted into the record of the proceeding and it included laboratory test information on both Digideck and WavePhore. The Electronic Industries Association (EIA) and NAB, on behalf of the NDBC submitted the laboratory data and the Commission approved use of the Digideck system.

Digideck's technology places a small carrier in the TV RF passband 1 MHz below the TV picture carrier and delivers a data rate of 525 kbps. A full system description of Digideck's technology is contained in Appendix B as part of the NDBC Final Report.

En Technology

En Technology's system encodes data in the VBI as well as over the visible portions or the TV picture (the data replaces the active video) transmitted by the television station. This system called *Cybercast* (originally *Malachi*) can deliver anywhere from 125.7 kbps (VBI) to 2.3 Mbps (full screen) to PC's equipped with a special *TV modem*. The modem can be programmed to permit reception of data in a variety of formats from VBI to any configuration of on-screen data. The on-screen data may be configured to start on any line of video, span any number of lines of video and any portion of each line of video, for applications where a mix of data and video is required.

Extended Data Service (XDS)

Extended Data Service (XDS) is a VBI data broadcasting scheme that allows broadcasters to send information about the station and its programming to consumers, such as the station's call sign, network affiliation, program title, length of the program and other information. The XDS data is interleaved with closed caption data on line 21 of the TV signal and its technical specification is documented within EIA Standard 608. XDS is also the technology that will likely be used to implement the V-chip blocking functions in TV receivers. The exact methodology will be determined pending the outcome of the FCC's TV ratings proceeding.

In May 1993, XDS was approved for use in the *Report and Order* in MM Docket No. 92-305 (which amended Rule sections 73.682 and 73.699 authorizing enhanced closed captions and other data services on field 2 of line 21).[1] Today there are only a handful of TV receiver models that respond to XDS, but that number is growing.

Microsoft/Yes! Entertainment

On October 24, 1996, the FCC approved Microsoft Corporation's data broadcasting technology. In a letter to Microsoft, the Commission stated that broadcast stations could use Microsoft's technology without prior authorization from the FCC. The Commission approved the use of this system because it is ostensibly the same as the technology developed by Yes! Entertainment Corporation that the Commission authorized in the *Report and Order* in MM docket No. 95-42. The Commission stated that the timing, amplitude and frequency parameters of Microsoft's technology were all within those of the Yes! system and thus should not pose problems for broadcasters.

Microsoft's system inserts data in the horizontal overscan area of the of the baseband television signal. It uses a one microsecond wide pulse (9.2–10.2 μs) placed between the end of the horizontal blanking interval and the beginning of the picture (the *back porch*) on each video line. The data is not placed on lines 1–24 of each field in order to avoid the VBI and any data being carried there. A data "1" symbol is represented by a luminance value of 80 IRE and a data "0" is represented by level of 7.5 IRE. The transmitted bit rate is approximately 14 kbps.

---

[1] This *Order* also reserved VBI line 19 for the ATSC Ghost Canceling Reference signal.

Teletext

In addition to closed captioning, another VBI data broadcasting service that has been in existence for some time is the *North American Basic Teletext Specification* (NABTS) or *Teletext.* Teletext is a technology that was developed in the 1970s as a method of delivering text information that could be displayed on a TV screen. It became very popular in Europe and Canada but never really caught on in the United States. NABTS specifies the use of multiple VBI lines and supports data rates up to approximately 56 kbps. Its technical specification is documented within EIA Standard 516. In recent years, NABTS has emerged as a general purpose VBI data delivery system because the technology is mature and very robust.

WavePhore, Incorporated

WavePhore was a participant in the NDBC evaluation program. Their TVT1/4 technology was approved for use by broadcasters in MM Docket No. 95-42. That proceeding was started partially because of a *Request for Declaratory Ruling* sent to the Commission by WavePhore. WavePhore's technology encodes baseband video in such a way that a subcarrier appears at the edge of the video passband in the RF signal. This system delivers a data rate of approximately 300 kbps. A complete system description is included in Appendix B as part of the NDBC Final Report.

National Data Broadcasting Committee (NDBC)

The NDBC was formed in 1993 and was co-sponsored by the National Association of Broadcasters and The Consumer Electronics Manufacturers Association (CEMA—a sector of the EIA). The Committee's purpose was to evaluate technologies that could be used for high-speed data broadcasting via over-the-air NTSC television transmissions. The NDBC held a total of ten meetings. The first was on September 16, 1993, and the last was held on December 17, 1996. The committee was co-chaired by a representative from both the broadcasting and consumer electronics industries. The co-chairs were Mr. Lynn Claudy, NAB and Mr. Werner Wedam, Sharp Laboratories of America representing CEMA.

The NDBC had two working groups—the Evaluation Working Group chaired by Mr. Bernard Caron (Communications Research Centre) and the Testing Working Group chaired by Mr. Victor Tawil (Association for Maximum Service Television). Originally, five companies submitted proposals for consideration in response to the committee's request for proposals but only two advanced to laboratory testing: WavePhore and Digideck.

Laboratory Testing

The committee evaluated the Digideck and WavePhore systems in laboratory tests during 1994 and 1995. The tests were performed at the Advanced Television Test Center (ATTC) in Alexandria, Virginia and were completed in October 1995. The laboratory program included the following tests that examined:

- Degradation to the host NTSC system
- Out of band emissions
- Impact on VBI services
- Interference
- The susceptibility of the data broadcasting systems to impairments

Numerous other tests were performed on the two candidate systems in order to observe issues such as recordability and real time delay. In June 1996, based on the results of the lab tests, the committee recommended both systems for evaluation through a field testing program.

Field Testing

The NDBC conducted field tests on Digideck's television data broadcasting system during the fall of 1996. (Just prior to beginning the tests, WavePhore elected not to participate in field test program.) The goal of the tests was to evaluate the performance of the data broadcasting system in over-the-air transmission conditions. The tests began on August 1, 1996 and were conducted in and around the Washington, DC metropolitan area.

The data signal was broadcast from two local stations: WJLA (Ch 7) and WETA (Ch 26). Measurements were taken at distances up to 45 miles away. Measurements at 100 outdoor locations were completed on August 30, 1996, and ten in-home sites were completed on September 18. The field test report indicated that satisfactory reception of data broadcast using Digideck's system at VHF was found at 93% of the test sites as compared to 82% satisfactory NTSC reception. At UHF, over 74% of locations had satisfactory data reception compared to 62% for NTSC.

The field test report was approved at the December 17, 1996, meeting and it stated that overall, the test results indicate that the Digideck's system provides data broadcasting service generally equivalent to the NTSC service area.

Final Actions

The NDBC voted at its final meeting on December 17, 1996, to create an *NDBC Final Report* that would include all of the technical work of the committee. In March 1997, that report was submitted to the FCC to become part of the Commission's technical record in Docket 95-42 on data broadcasting services. For a copy of the report contact NAB Science and Technology at (202) 429-5346.

ATSC DATACASTING

In addition to high definition video and multichannel audio, the new ATSC television standards support data broadcasting or *datacasting.* Datacasting may have just as much impact on the success of digital television as the audio and video.

ATSC datacasting services delivered at multi-megabit per second speeds and new intelligent receivers have the potential to make digital TV (DTV) far more valuable to the consumer and give broadcasters an opportunity to offer new services, generate additional revenue and accelerate the consumer migration to the new broadcast standard. The challenge for broadcasters is to begin to develop effective data content and services to support the next round of intelligent ATSC devices.

Two Datacasting Categories

ATSC datacasting applications are generally grouped in two categories:

- **Enhanced Television.** Data content related and synchronized to video program content.
- **Data Broadcast.** Data services not related to program content. An example would be current traffic conditions, stock market activity or even subscription services that utilize ATSC conditional access capabilities. Data broadcast bandwidth could also be sold to third parties for a wide range of applications not related to broadcast television.

This chapter will focus primarily on the enhanced television applications for ATSC datacasting applications that could redefine consumer expectations for television. It will describe a range of likely applications that could be supported by the proposed ATSC standards and the intelligent ATSC receivers and PC based receivers that manufacturers are developing.

There are a number of ways the ATSC datacasting capabilities could be used to enhance program content and provide a compelling reason for consumers to purchase an ATSC converter box for their existing television—or even a complete digital television.

Based on the specifications detailed in both proposed standards, and the capabilities that the consumer electronics companies are considering for their ATSC receivers, what follows is a description of what a viewer might see within next few years if they have an intelligent ATSC receiver.

The viewer is watching a travel program. When there is datacasting content related to the program, an icon could appear on the screen in an unobtrusive manner (see Figure 6.4-1). If interested in seeing more information, the viewer could push a button on the remote.

Figure 6.4-2. ATSC Screen with second level menus.

The next level of information, also unobtrusive, could identify the types of content available (see Figure 6.4-2).

If a category looks enticing, the viewer could select the category—for example *travel*, then choose whether to the view information immediately or bookmark it and retrieve it later (see Figure 6.4-3). Bookmarking would instruct the receiver to store the data in local memory for later use.

If the viewer decided to look at the datacasting content immediately, the moving video image could shrink and reveal additional menus allowing the viewer to go a level or two deeper for information (see Figure 6.4-4). Again, the viewer could bookmark it (store it for later viewing), view it on screen immediately or even print it.

Another feature being considered by many receiver manufacturers is a built-in reverse communications channel (such as a modem) and Internet browser. The ATSC standard will likely support links to a related website, allowing receivers with Web browsers to automatically connect to a website to support interactive applications, even ordering goods and services.

If there is a reverse communications channel, the viewer could request additional information, generate an automatic e-mail or transparently link to an Internet site.

Notice the ad in the lower right side of the screen in Figure 6.4-4. Datacasting can provide new opportunities for advertising, in this case a resort hotel in the area.

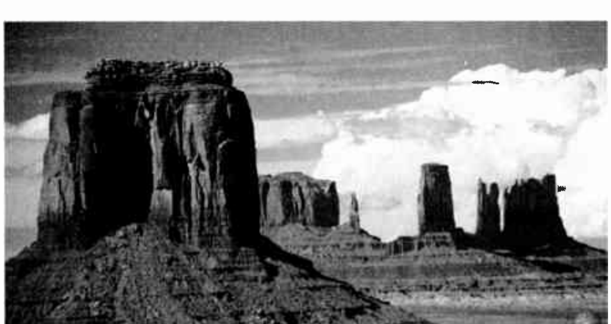


Figure 6.4-1. ATSC display with datacast icon.

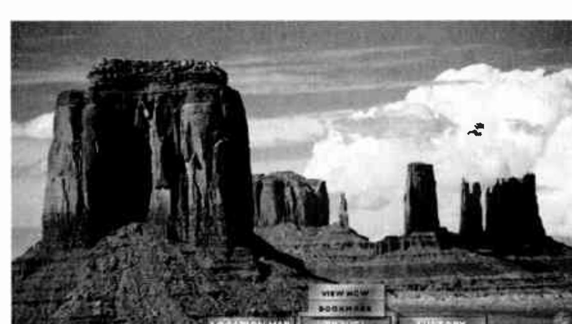


Figure 6.4-3. ATSC screen with *bookmark view now* menus.

Figure 6.4-4. ATSC screen with reduced picture area and hotel ad.

Figure 6.4-6. Football game screen.

Distracting the Viewer

Producers, broadcasters and most import advertisers have voiced concerns about distracting a viewer with data content during the program and the possibility of moving the viewer's attention away from the television program to the Internet. These are valid concerns with no clear answers. Nevertheless broadcasters, producers and advertisers are already developing scenarios for enhancing program content with datacasting.

For advertising, one technique with great potential is expanded use of placement promotions—a viewer watching anything from a situation comedy to a home improvement program could have immediate access to anything shown in the program. If a viewer liked the clothing, the furniture or the car, with the push of a button more detailed information could be available immediately. The value of product placements in home improvement shows could increase significantly if a viewer could obtain more information about the product and where to purchase it.

For conventional advertising, datacasting could give viewers the opportunity to respond immediately if they wanted more information. In Figure 6.4-5 information about biking events, products and where-to-buy are available. Datacasting could also support a second level of advertising enhancement. Once the viewer has responded to the primary message, datacasting gives the advertiser an opportunity to promote related products. In this ad for bicycles, a second level menu can also promote helmets and clothing. If the receiver supports a reverse communications channel, the viewer could respond directly to the advertiser, register for a related biking event, or even purchase a product.

In a sporting event, when the datacasting icon appears, the viewer could push a button on the remote to bring up the next level of menus and select a category, for example, merchandise, and decide to look at it immediately (see Figure 6.4-6). The viewer could select a team to see what is available and even purchase it, all the while watching the game.

There is enough data carrying capacity in the ATSC standard to download information to support multiple menus—every possible menu selection in this example could easily be downloaded in less than 10 seconds. Of course the receiver would have to be equipped with enough memory to store the various menus. The proposed standard will likely have a provision to carousel data, that is broadcast it multiple times, making it available to support a viewer's selection within seconds even if the receiver does not have sufficient memory to store all of the data necessary to support a series of menu options. There is also enough broadcast data capacity to send pictures and multimedia content—video and audio—and use new techniques such as immersive photography and JAVA applets.

Datacasting could also enhance news programs. For this example (Figure 6.4-7), video content in a 4:3 aspect ratio is used allowing the news program to be simultaneously broadcast for NTSC and ATSC viewers. But the ATSC viewers would have immediate access to additional information such as transcripts, still images and related stories. A digital viewer could

Figure 6.4-5. Diablo bike ad.

Figure 6.4-7. News screen.

also have access to local information such as weather, traffic and more.

Even with the same video program content, this scenario demonstrates how ATSC datacasting can provide benefits to consumers who simply purchase a relatively low-cost ATSC converter box capable of supporting datacasting.

Of course, a broadcaster could use datacasting to provide local information such as weather and traffic 24 hours a day, making the television in the family room a primary source for instant information. Broadcasters might also offer specialized subscription services that could entice select groups of viewers to tune to their station for specialized information. For example a broadcaster could add doctors, financial professionals and other select demographic groups to its audience because of the unique subscription services it offers.

Standards for Intelligent Receivers

Consumer electronics and PC manufacturers are actively designing a new generation of flexible DTV receivers capable of processing and displaying datacasting content. There are three standards. The fundamental one way data transport, the protocol for a return channel to enable interactive services and receiver environment standard. The receiver environment standard defines the processing software and application program interface routines.

The standards allow content to be created in a single format that would automatically be presented in alternative formats depending on the receiver's capabilities integrated very well with the Application Execution Engine and Java framework making them relatively platform independent.

Existing Web authoring tools should be usable by content development as well as support reuse of most existing Web content. Reformat of content datacasting applications and TV display can be performed. Existing transaction and order fulfillment servers used for Web-based services are supported.

Links between a program related website and a broadcast program are supported. A viewer can interact with the content included in the datacasting stream or a broadcaster could easily create links between the broadcast program and a program related website.

Support for Web browsing in order to support useful datacasting services is not required. Low cost receivers will be able to process and display datacasting content without requiring a browser and reverse communications channel.

Program related content is supported with techniques to synchronize data content to specific segments of a video stream and precise layout control for data content to coexist with a video image.

DTV Receivers

DTV receiver manufacturers have several options to develop receivers capable of handling datacasting. One is the traditional PC architecture.

An alternative approach taken by several consumer electronics companies is to use low cost programmable media processors. These processors handle ATSC stream demultiplexing, decoding, plus control and communications functions on a single chip. Some of these media processor chips can even support a complete Internet browser, software modem and video e-mail.

It is clear that consumers will soon expect digital television receivers to bring more than just TV into their homes. ATSC datacasting already has producers developing enhanced content for their programs, advertisers and consumer products companies testing how to use datacasting effectively and the consumer electronics and PC companies developing the next generation of intelligent, programmable DTV receivers to bring new services into viewers' homes.

For further information, visit:

http://www.toocan.philabs.research.philips.com/misc/atsc/dase

http://www.toocan.philabs.research.philips.com/misc/atsc/bhtml

http://www.atvef.com

6.5
TRANSMISSION SYSTEM CONTROL AND MONITORING

JOHN E. LEONARD, JR., CONSULTANT
GENTNER COMMUNICATIONS CORPORATION, SALT LAKE CITY, UT

INTRODUCTION

Transmission system control and monitoring is often simply perceived simply as transmitter remote control. However, it must also be recognized that, unlike radio, television is more station specific with regards to transmitting equipment and how it must operate. While radio has AM and FM, television has VHF, UHF, lower power stations, boosters and translators. Translators exist in FM, but television has typically regarded them in a different manner. Before exploring what is now taking place, a brief review of the current regulations for transmitter remote control is appropriate.

In 1953, the Federal Communications Commission (FCC), under Part 73 of its Rules and Regulations, permitted remote control of broadcast transmitters. Initially AM radio stations, up to a given power level, were authorized to operate via remote control. By 1957, most radio stations were allowed to operate with remote control. The first television transmitter remote control operation was permitted in 1963. At that time only UHF stations could use transmitter remote control. In 1971, VHF finally had the ability of remote controlling their transmitters. Also at that time, all classes of broadcast transmitters had remote control capability available in some way.

In 1977, the radio rules underwent some changes, and in 1984 the Rules were amended to have remote control under a uniform code; in addition, there was a relaxation of certain previous requirements.

The most recent Rules changes occurred in late 1995 and impacted full power television (Part 73) and all other over-the-air television (Part 74). The FCC action under MM Docket No. 94-130 (FCC 95-412) recognized unattended operation of broadcast stations and it updated transmitter control and monitoring requirements. These regulatory requirement changes within Part 73, specifically for television, are the focus of this chapter. The changes in Part 74 relate to low power television and translators and simply mirrored those requirements within Part 73. Current industry practices and typical equipment used will also be discussed.

Digital television (DTV) was officially introduced to the industry in 1996 through on-air field tests. DTV is addressed in this chapter, but keep in mind that DTV was in the early stages of implementation at this writing. Other chapters of this *Handbook* covering related subjects will also be cross-referenced here.

APPLICABLE REGULATIONS

UHF television began using transmitter remote control in 1963 and VHF in 1971. The 1984 Rules amendments placed remote control under a uniform code and relaxed some previous requirements. The changes that occurred in 1995 from the FCC MM Docket No. 94-130 (FCC 95-412) affected all areas of broadcast. Part 73 will receive principal attention here. Specifically, unattended operation and *automatic transmission systems* (ATS) were among the areas addressed. FCC 95-412 created Parts 73.1300 and 73.1350 and revised Part 73.1400. These sections have become the principal controlling parts of the FCC Rules and Regulations in relation to transmitter remote control.

It is interesting to note the parts that were removed by FCC 95-412. Sections 73.1410 *Remote Control Operation*, 73.1500 *Automatic Transmission System* and 73.1550 *Extension Meters* were all deleted.

The *Discussion, Universal Application* section of FCC 95-412, Paragraph 10, contained the following:

> Therefore, all types of standard (AM, FM and TV) stations, as well as international broadcast stations will be permitted to operate unattended. The same permission applies to low power TV (LPTV) with locally originated programming.

Also involved was the implementation of Emergency Alert System (EAS). At the time of the adoption of FCC 95-412, EAS was still a new concept. (EAS must be observed when operating via remote control, attended or unattended.) Chapter 1.12, *EAS Standard* presents detailed information about EAS.

The specific permissible methods of remote transmitter control added by FCC 95-412 were the use of the public switched telephone network in lieu of the previous dedicated interconnecting circuits or lines. While it removed Section 73.1500 (Automatic Transmission System) this Docket redefined ATS and how it can be used. In Section 73.1580 the Docket also clarified the transmission inspection requirements to be *periodic*. The clarification of ATS is found in 73.1350 and 73.1400. Also clarified in Section 73.1635, *Special Temporary Authorizations* (STA), was the relationship to Section 73.1350 (and the interrelationship to other sections).

In general, remote operation of any broadcast transmitter is under the control and the responsibility of the licensee. In the *Discussion, Unattended Operation* section of FCC 95-412, Paragraph 7 was added:

> 7. The Commission, based upon its experience in enforcing broadcast rules, concurs with the majority opinion that waiver of Section 318 of the Act to permit unattended operation is not likely to result in an increase in operation outside the tolerances specified in the Rules or the station authorization and will not adversely affect the public interest. Significant technical malfunctions should quickly become obvious and we do not anticipate their continuance for significant periods of time. The waiver appears further justified for reasons of efficiency, in order that our broadcast licensees can best decide how to allocate resources to ensure compliance. Therefore, the rules relating to station operators will be eliminated or revised as proposed in the Notice so as to permit unattended broadcast station operation.

In essence the prevailing concept from FCC 95-412 is broadcasters should do what is best for their station in relation to remote control. But it also indicates that each station must be sure that what it is doing is appropriate. No longer has the Commission attempted to provide specific requirements in lengthy Rules and Regulations. Rather, the sections referenced by the current Rules simply need to be understood in for each station's precise situation.

With DTV coming into being, there were specific section(s) added to the Rules and Regulations to accommodate digital transmitters. While digital television transmitters are clearly different from existing analog transmitters, the actual operating theories remain such that the current applicable sections will apply.

Sections 73.1300, 73.1350 and 73.1400 are reproduced at the end of this chapter in Appendix A. Section 73.691, *Visual Modulation Monitoring* is also of importance, as are 73.1820 *Station Log* and 73.1870 *Chief Operators.*

CURRENT PRACTICES

Transmitter remote control came into existence first to keep stations on the air and second to do so while saving money. Remote control was not intended to alter the method of operation for a given television station. Remote control must keep the station on the air, and monitor what is occurring. This application becomes a lofty requirement for remote control with ATS-type possibilities and exacting modulation monitoring needs.

Radio has its own unique requirements for remote control as does television (see Chapter 4.9 *AM Antenna Coupling and Phasing Systems*). With television (existing, analog transmitters) the first technical difference relates to the fact that a TV transmitter is actually two transmitters—one FM in nature (the aural transmitter) and the other AM in nature (the video transmitter). Basic technical differences exist for the monitoring requirements of each. The specifics of each transmitter determine what parameters must be monitored.

Modulation monitoring of the transmitter must occur. Section 73.691 outlines the visual modulation monitoring requirements and references 73.682. 73.682 provide the particulars of both the analog visual and the aural transmitters. It is common practice to separate the actual signal measurement and monitoring from the transmitter remote control system. Some stations, however, do interrelate transmitter remote control with signal measurement and modulation monitors.

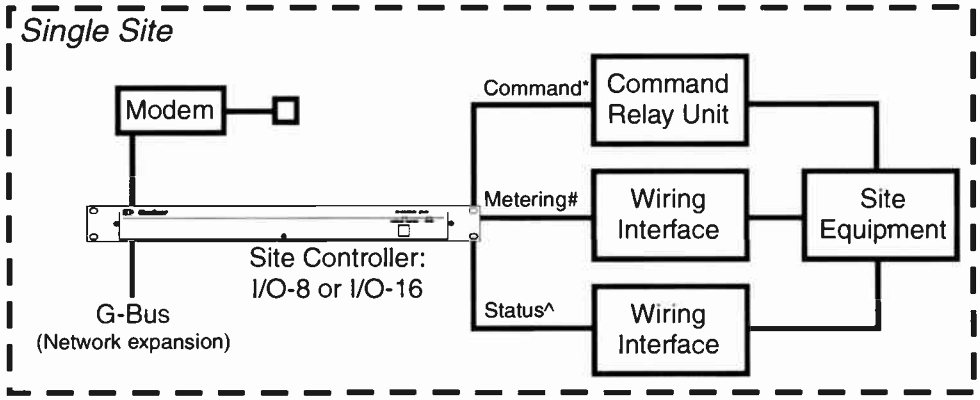
As is evident from an examination of the Rules, the actual delineation is left undetermined. This enables the interrelationship of monitoring equipment and transmitter remote control equipment, making ATS—and similar—possibilities almost unlimited.

Transmitter remote control equipment has had several classic characteristics that relate specifically to the transmitter. First, is the ability to present or display actual parameter values in real-time (metering). Second is the ability to issue commands—the ability to actually control something. Finally, a simple go/no-go display or indication is provided (status). The requirements of specific transmitters have made the implementation of these functions unique to broadcasting. A general overview of transmitter remote control functions is given in Figure 6.5-1.

Metering has unique requirements created by the parameter being monitored. For example, the measurement of transmitter RF power output is of significant importance. To ensure that the transmitter remote control system is presenting the correct numeric value, calibration or verification of power output is necessary. One factor that impacts this parameter is the method used to sample power. Because the reflectometer is the standard device used for power measurement, the mathematical relationship of the sampled value to the value delivered to the remote metering system is critical. Many transmitter remote control systems include various modes of operation for each metering channel. With this equipment being microprocessor based, power-to-linear conversion is practical for multiple metered channels.

Some will also choose to use the classic indirect method of determining power output, where the input parameters of the amplifier are multiplied and a mathematical correction is applied based upon the actual efficiency of the specific amplifier. Transmitters that produce a sample of power output that is linear in nature also are available thus making this interface easier.

Recent Rules revisions have fostered an interrelationship between the capabilities of the modulation monitoring equipment and the remote control equipment. This interconnection concept makes possible



* Either a solid-state type or dry-contact closure output. Typically this type of equipment has these as a joined pair (two outputs) and referred to as raise/lower. This was due to the original need to control the output power of the transmitter.

Commonly referred to as a DC sampling voltage. This DC voltage has a direct relationship to the actual transmitter parameter. These are typically in the range of 0.5 to 3.5 Vdc.

^ Typically a single or double-state condition, such as a dry-contact closure or it can be in a solid-state or digital condition or state. This digital state further simplifies the direct interface between this and other monitoring equipment.

Figure 6.5-1. Typical transmitter remote control equipment input/output configurations.

unique ATS implementations that are specific to a single transmitter site.

Modern transmitters provide what their manufacturers view as required elements for proper operation. Power output, at the proper level, may be the prime area of concern for a station. A review of the *FCC TV Broadcast Station Self-Inspection Checklist* (Bulletin CIB-18TV) confirms the emphasis on control and metering of output power from the transmitter. To be understood from that review is the need for maintaining and monitoring power output, turning the transmitter on and off, monitoring modulation and observing the functionality of tower lights.

Input and Output Functions

The primary function of a transmitter remote control system is monitoring and controlling the transmitter. With the emerging need to monitor more than just the basic transmitter parameters, both metering and status inputs may need to interface with other equipment, such as modulation monitors. The specifics of such interfaces are determined by the actual specifications of the equipment involved. As an example, it could be useful to take an over modulation output signal from a modulation monitor and interface it with a status input. The status inputs could also be interfaced with security systems and other ancillary devices. A similar capability exists with metering inputs.

Remote control equipment manufacturers provide sampling devices for a variety of parameters. Many of these devices relate to the metering input functions. Examples include temperature, ac line voltage and current. Such sensors can be used to observe of the operating parameters of tower lights (if not equipped with their own sensors), air conditioning equipment and other critical systems.

The command output interface may assume one of many forms. As is noted in Figure 6.5-1, both solid-state and dry-contact closure command outputs are useful for transmitter remote control. The dry-contact (relay) interface is usually the most flexible, although it is also usually the most expensive. A similar situation can occur with status inputs. The specific requirements of some status inputs may dictate an interface device. Specific solutions in this area are typically provided by the remote control equipment manufacturer.

Transmitter Interface

Modern television transmitters—as a general rule—include all of the remote control interfaces required for normal applications. As such, connection of the transmitter to the remote control system is a comparatively simple task. The actual physical locations of this equipment may present more of a challenge than the electrical specifications.

Attention must, however, be paid to the exact characteristics of the sample provided by the transmitter and the input needs of transmitter remote control equipment. For metering interface to the remote control equipment, some transmitters may require a dc isolation amplifier or similar device. The same caution applies to the status input and command output interface. Verifying all specifications when planning a new system is advisable. Extensive discussion of interface methods and devices is given in Chapter 4.7, *Transmission System Control and Monitoring*.

Automatic Transmitter System Operations (ATS)

As stated previously, an ATS system can be established by the station licensee. Such operation may, in fact, be provided from the combination of a number of separate systems. Depending upon the class of license and its related transmitter, ATS may be performed totally by the transmitter itself. Such capabilities can also be combined with those of the transmitter remote control system to enhance the overall capabilities.

Modern transmitter remote control equipment has the capability of comparing actual metering inputs with preset values. When those values are exceeded, the remote control equipment can initiate a correcting action, or simply sound some form of alert to report that the condition has happened and exists (or existed). The same function also is common with status inputs. Some tower lighting control systems have similar capabilities. When these multiple capabilities are then combined with other smart devices and systems, the ATS capabilities can become almost unlimited.

Automatic Transmitter Logging Systems

Certain transmitter remote control systems have the ability to meet not only ATS requirements, but also some of the recording capabilities required by the Rules. In the past this function was referred to as *automatic logging*. Some existing transmitter remote control equipment can provide at least segments of what the current Rules require, given the interconnection possibilities presented. Considering the calibration confirmation requirements in the current Rules, it is the author's interpretation that a totally automatic logging system for a station may not be currently economically feasible. Recording the calibration of all the systems involved in an ATS system may go beyond the capabilities readily available. Still, certain elements are in place today that fulfill many of the stated requirements.

For example, Figure 6.5-2 is a screen capture of a computer based logging (recording) transmitter remote

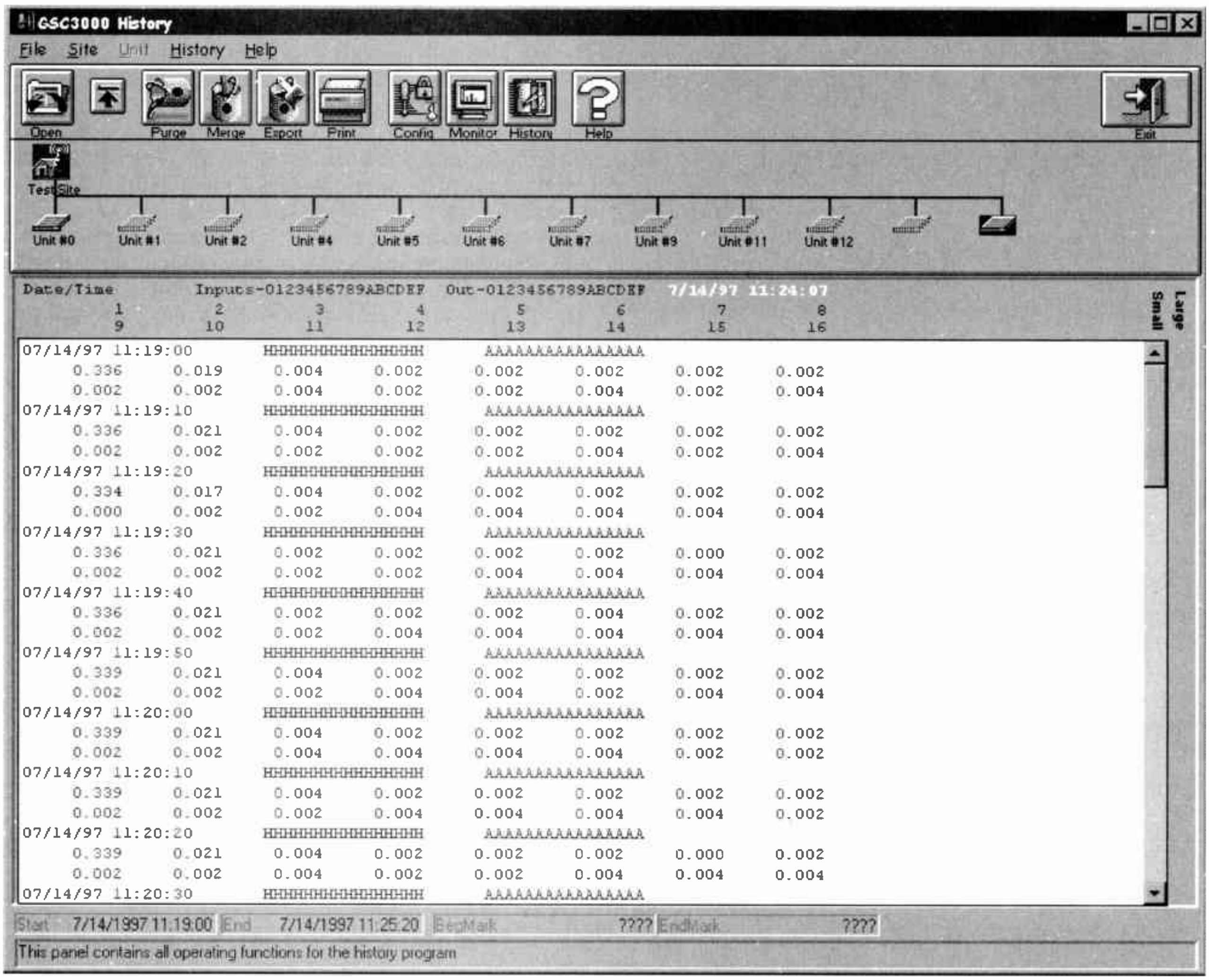


Figure 6.5-2. Data stored within a PC Windows® presentation from networked transmitter remote control equipment.

control system. Such logging can meet the majority of the determined needs. Again, preplanning for installation of the transmitter remote control can be of great help.

CURRENTLY AVAILABLE EQUIPMENT

Transmitter remote control equipment has historically been classified by the number of *channels*. These channels relate to metering capability, status capability and command capability. Some noticeable changes have recently occurred in the equipment specifically supplied for remote television transmitter operation. As the word *digital* became prominent in television in general, the same technology found its way into transmitter remote control equipment. The addition of networking capabilities to this type of equipment is having and will continue to have a significant impact. Networking adds expansion capabilities to the numbers—number of channels and number of sites.

In a classic remote control system, information is simply moved from one place to another, usually from the transmitter site to the studio. The information moved is the remote display of metered values, such as plate or emitter voltage and currents, plus power output. Status is used for indicated situations. Included also is the issuing of commands back to the site. These functions have been accomplished in a digital manner for some time. The 1984 Rules change added capabilities that some in the industry had clearly favored. Those basic capabilities were the use of standard dial telephone circuits and the use of touch-tone command capability with voice response for information feedback. The equipment that provided those capabilities was microprocessor based and also had internal decision making capabilities. Those capabilities are what then permitted unattended operation without human supervision, resulting in the 1995 Rules changes.

The equipment that is available today is all digital; many—perhaps most—are microprocessor based. This creates an environment in which the total system can have the ability to be smart. By smart the implication made is the equipment, on its own, makes a decision or takes action. Advanced transmitters also include such capabilities, as do some modulation monitoring equipment. Thus, by combining the transmitter remote control equipment with a smart modulation monitor (and a smart transmitter, if so equipped) the transmitter site location can thus become an ATS system.

In the original physical conception of a transmitter remote control system, two units were used—the transmitter and the studio. Predictably, one was physically located at the studio, and the other was at the transmitter site. Television requirements, as compared with radio applications, can dictate a massive transmitter remote control requirement. Each station has its own particular situation. Many have two or more transmitters—a main and backup or standby. Additional transmitter sites also can exist, such as for translators and booster transmitters. A television station, thus, can have many transmitters at many different locations.

Communication between these distant points typically occurs in some analog or digital form. One analog form that entered the industry with the 1984 Rules revisions was the use of a telephone instrument as the studio unit of the system. A touch-tone (DTMF) type instrument became one end, with the transmitter unit accepting DTMF codes for commands and producing a synthesized voice for the returning data. This type of unit used a microprocessor (thus it had smarts). While television typically has used this concept, it was not until 1995 that this approach became the principal method of television transmitter remote control. Other digital systems were used, with the other end typically being either a digital studio unit or a personal computer (PC).

Attributes and options for current remote control systems include the following:

- Studio/transmitter unit, with PC monitoring (at either or both locations): interconnection via a dedicated telephone line or interconnection via a dedicated wireless link
- Transmitter unit, with personal computer monitoring: interconnection locally, at transmitter, between PC and unit; interconnection via a dedicated telephone line; interconnection via a dedicated or shared wireless link; or interconnection via the public switched telephone network (PSTN).

A *dedicated wireless link* typically refers to the use of sound channels on a two-way video studio-to-transmitter link, or a sound channel to the transmitter site and a subcarrier on the aural transmitter for return data or data within the video. The *wireless links* using PSTN and plain old telephone system (POTS) interconnections refer to cellular or similar systems. A graphic presentation is given in Figure 6.5-3.

The PC has been linked to transmitter remote control for some time. The various operating systems of the PC have garnered much attention in the last few years. DOS was supplemented by Windows, then Windows moved to Windows 95, and to Windows NT. Figure 6.5-4 shows a Windows based remote control display. The graphic presentations provided by the operating system support multiple ways of presenting the conventional metering, status and command information. In Figure 6.5-4, metering is presented numerically as well as by bar graph.

Networking transmitter remote control has taken a major step forward by combining the strengths of the dedicated transmitter remote control system with the PC. The use of PC based smarts provides a number of user benefits, including:

- An accurate time base (time-of-day and date)
- Reliable and flexible supervisory command and data acquisition capability
- A near infallible memory capability for operation and recording data
- Addressability and changeability
- Multiple communications capabilities and types
- Off-the-shelf manufacturing economies of scale.

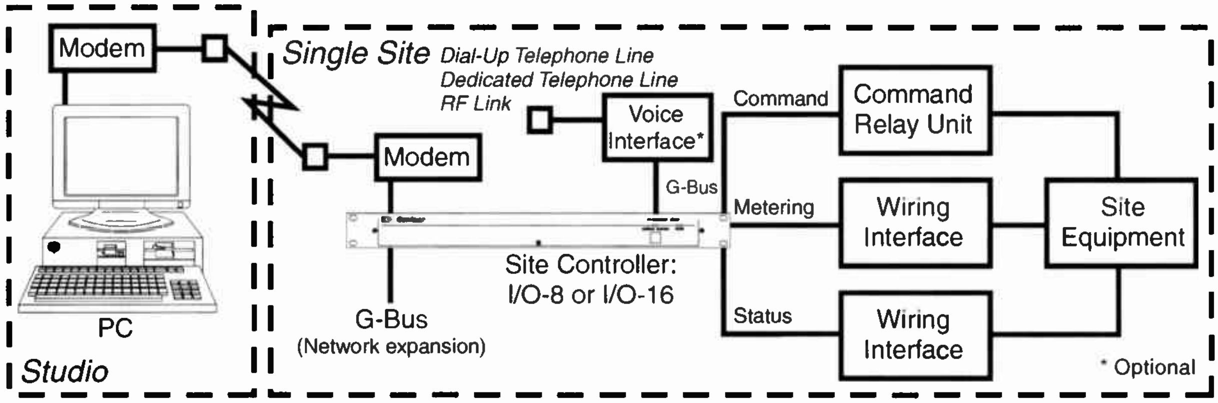


Figure 6.5-3. Possible interconnecting circuits for transmitter remote control equipment.

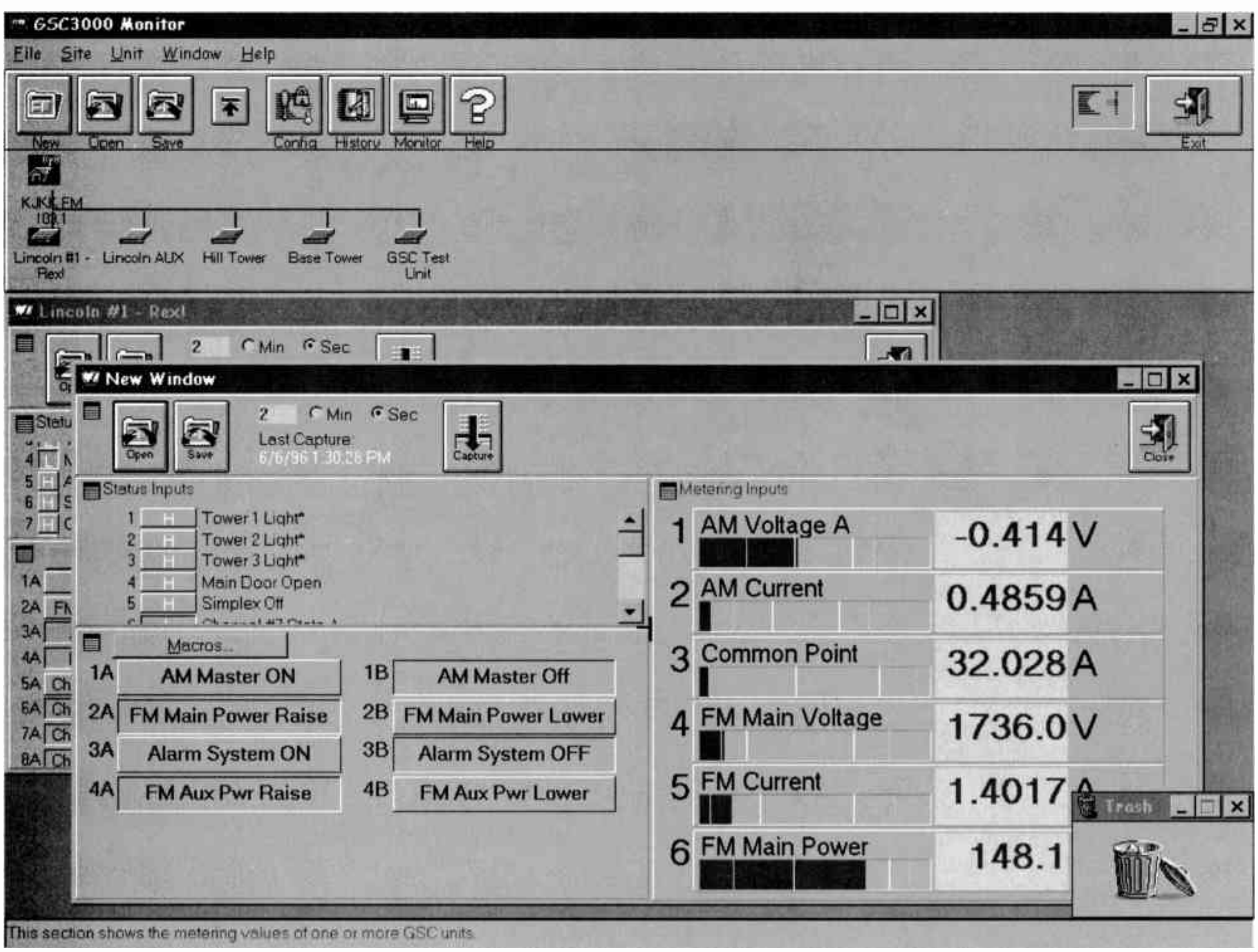


Figure 6.5-4. PC Windows® presentation from networked transmitter remote control equipment.

Networking includes more than just selection of the type of interconnecting circuit to be used. Rather, it enables an enormous expansion of capabilities. With the support of a controlling processor, a number of additional capabilities can be added, including:

- Limit monitoring and checking against preset points at multiple levels for each metering channel
- Selection, from each preset point, whether the measured value has created what can be classified as an alarm
- Selection, from each preset point, whether the measured value has crossed the point at which a command output should be activated
- The capability, beyond issuance of a command, to activate a chain of preprogrammed events—in other words, a macro
- Accomplish, from a status channel, with a change of state, the three previous points (alarm, issue command or start macro)
- Use of the internal time base (clock/calendar) within the unit to issue a command or to activate an internal macro at a preselected moment
- Recording the occurrence of any of the previous set points within its non-volatile memory for subsequent examination and reporting
- Establishment of an interlink with one or more other local unit.

These are just some of the internal capabilities that can exist within a computer based remote control system. The interconnection capabilities, given the networking concept, go beyond the type of circuit needed at the transmitter site. Because the unit is processor based, expansion is possible. This expansion can occur on a network and between units. Each unit also has its own unique channel capability. Processors are best used when with another processor—thereby creating multiple networks. (Figure 6.5-5 presents some of these capabilities.) While each unit shown in the Figure is a separate entity, they combine to make a system that is greater than the sum of its component parts. A common protocol enables communications at the local area network (LAN) level. The LAN can be open and unrestricted or manufacturer unique and proprietary.

The individual ports shown in the Figure 6.5-5 system support outward communications, such as a hardwired LAN or a modem. Local communications, being

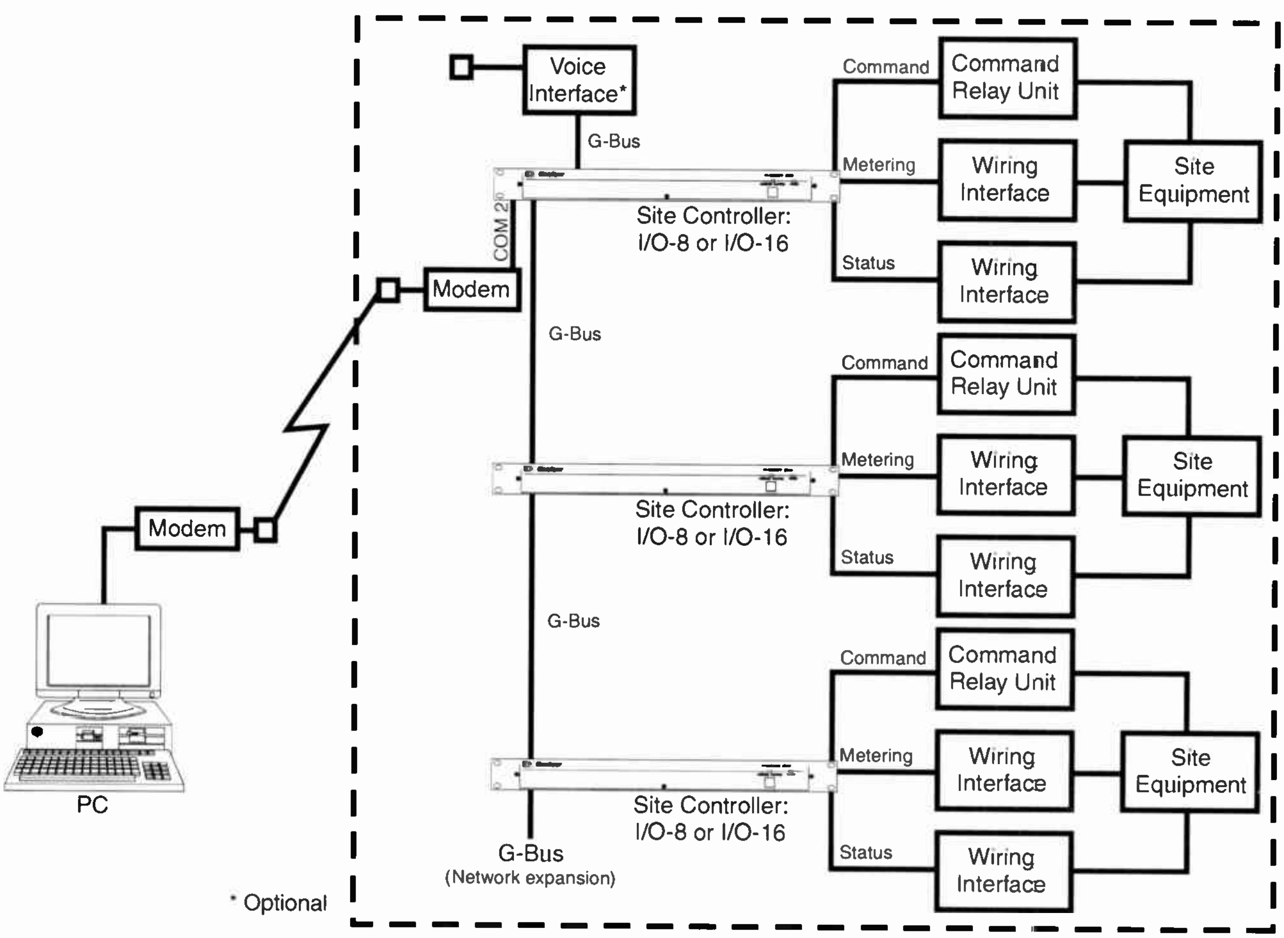


Figure 6.5-5. Networks used for transmitter remote control equipment.

pure data, can occur directly with a PC at the site. Multiple ports permit the unit to communicate not only with a local PC, but with a remotely located PC via a standard modem or other link. Through the use of a standard modem, the need for actual interconnection circuits can be simplified. Both 2-wire and 4-wire modems exist, thus supporting most commonly available analog communication circuits.

Figure 6.5-6 shows an actual transmitter status log. Note, at the top of the Figure is the date, the site and the time. The *header* then follows, with licensee, business address, call letters, channel and frequency, the remote control address, the transmitter address and the specific time being presented in the record. Next, are the actual entries selected, in a column format. Next is the measured data at the prescribed intervals, in this case, the data is taken every hour.

Further note the station's decision to have operators sign the printed log. This provides the station with an independent record of the individual they have placed in the transmitter operator position and the times that individual has been present. Other stations have made a conscious decision to have this capability, let it record, but have their operators maintain a handwritten log.

The network based remote control system is further refined in the configurations shown in Figures 6.5-7–6.5-9. Figure 6.5-7 illustrates the arrangement for one transmitter site and one studio. Figure 6.5-8 shows a site with expanded channel needs. The use of the POTS interconnection is shown. Figure 6.5-9 demonstrates how the network concept can expand as the number of transmitter sites increases. A voice interface may also be used to further enhance flexibility. When the PC and voice interface capabilities are combined, a system can be created to meet user specifications and needs.

For EAS applications, some interesting uses of available remote control equipment have appeared at some television stations. EAS equipment has the ability to digitally store the audio message delivered to it as part of an alert. The presence of that audio message is then sensed by a DTMF/synthesized voice transmitter remote control unit. The remote control can also relay the actual audio to the individual accessing it.

CONCLUSIONS

The FCC, with MM Docket No. 94-130 (FCC 95-412), created the opportunity for new technologies to be used in broadcasting. Without this docket, it is questionable whether transmitter remote control would have received the attention it deserved. FCC 95-412 motivated the development of networking capabilities within transmitter remote control, and the ATS potentials that now exist are unprecedented. This ATS capability may—in fact—prove very advantageous for television transmitter operation. These capabilities, and the new features that they offer, are now under the authority and control of the licensee. The new technologies, thus, can be used in the most efficient way for individual television stations.

```
Site 2  Wed 05/21/97 23:55:18
WGN Continental Broadcasting Company  2501 W Bradley Pl  Chicago, Illinois 60618
WGN-TV  Channel 9   Visual Frequency, 187:26 MHZ    Aural Frequency, 191:76 MHZ
Remote Control Address-2501 W Bradley Pl  Transmitter Address-875 N Michigan AvZ
                            All Times Central Daylight

Site 2  Wed 05/21/97 23:56:18
             A Aur    A Vis    B Aur    B Vis
             Power    Power    Power    Power
               %        %        %        %

Site 2      +  94.  + 100.
Thr 05/22/97 00:00:16


Site 2      +  92.  +  94.
Thr 05/22/97 01:00:17
```

Figure 6.5-6. An actual log taken by WGN television, using Networking Transmitter Remote Control Equipment. This log printed on printing device of associated PC, as well as displayed on PC monitor.

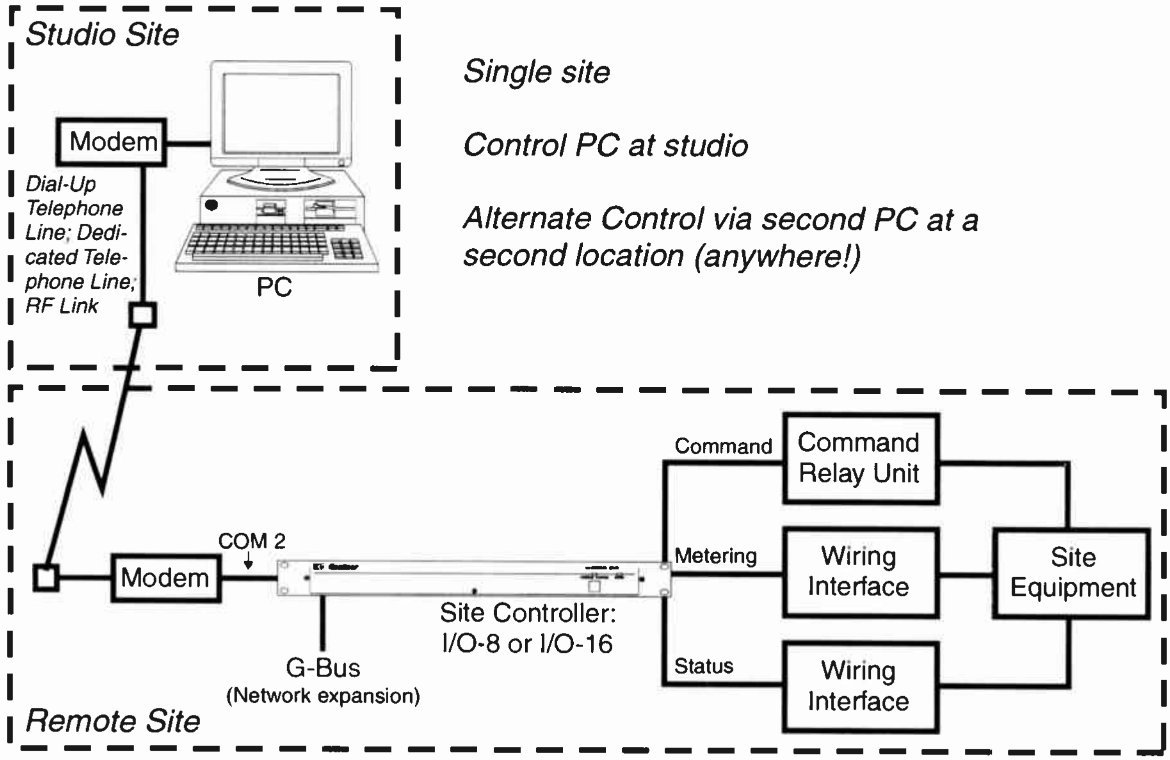


Figure 6.5-7. One site of networking-type transmitter remote control equipment using POTS interconnection.

APPENDIX A

FCC Regulations Parts—73.691, 73.1300, 73.1350, 73.1400, 73.1580, 73.1635, 73.1820, & 73.1870

§ Sec. 73.691 Visual Modulation Monitoring.

(a) Each TV station must have measuring equipment for determining that the transmitted visual signal conforms to the provisions of this subpart. The licensee shall decide the monitoring and measurement methods or procedures for indicating and controlling the visual signal.

(b) In the event technical problems make it impossible to operate in accordance with the timing and carrier level tolerance requirements of Sec. 73.682 (a)(9)(i), (a)(9)(ii), (a)(12), (a)(13), and (a)(17), a TV broadcast station may operate at variance for a period of not more than 30 days without specific authority from the FCC: provided that, the date and time of the initial out-of-tolerance condition has been entered in the station log. If the operation at variance will exceed 10 consecutive days, a notification must be sent to the FCC in Washington, D.C., not later than the 10th day of such operation. In the event normal operation is resumed prior to the end of the 30 day period, the licensee must notify the FCC upon restoration of normal operation. If causes beyond the control of the licensee prevent restoration of normal operation within 30 days, a written request must be made to the FCC in Washington, D.C., no later than the 30th day for such additional time as may be necessary.

§ Sec. 73.1300 Unattended Station Operation.

Broadcast stations may be operated as either attended (where a designated person is responsible for the proper operation of the transmitting apparatus either at the transmitter site, a remote control point or an ATS control point) or unattended (where highly stable equipment or automated monitoring of station operating parameters is employed). No prior FCC approval is required to operate a station in the unattended mode. Regardless of which method of station operation is employed, licensees must employ procedures which will ensure compliance with Part 11 of this chapter, the rules governing the Emergency Alert System (EAS).

§ Sec. 73.1350 Transmission System Operation.

(a) Each licensee is responsible for maintaining and operating its broadcast station in a manner which complies with the technical rules set forth elsewhere in this part and in accordance with the terms of the station authorization.

(b) The licensee must designate a chief operator in accordance with Sec. 73.1870. The licensee may designate one or more technically competent persons

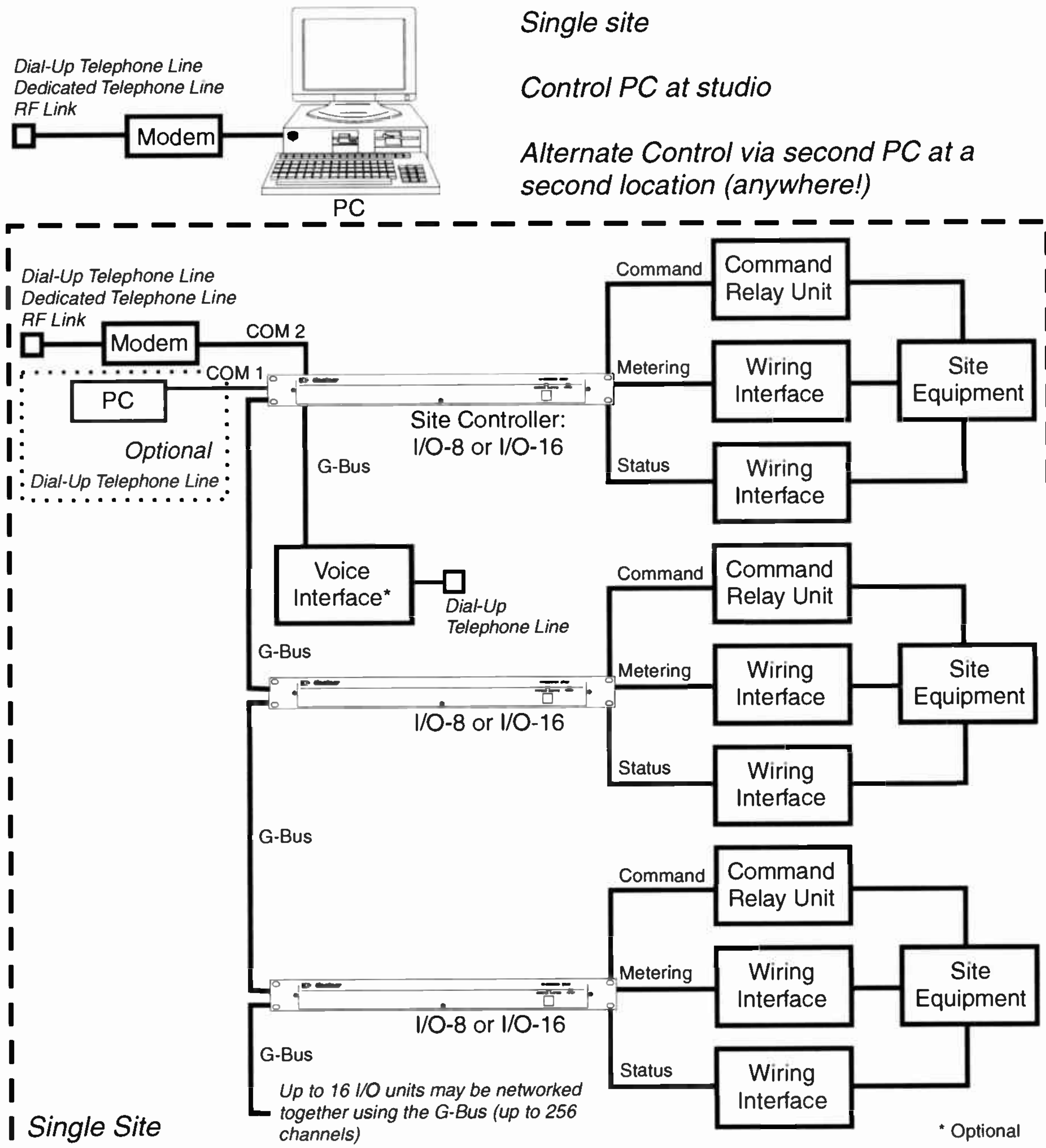


Figure 6.5-8. Another site of networking-type transmitter remote control equipment using POTS interconnection.

to adjust the transmitter operating parameters for compliance with the technical rules and the station authorization.

(1) Persons so authorized by the licensee may make such adjustments directly at the transmitter site or by using control equipment at an off-site location.

(2) The transmitter control personnel must have the capability to turn the transmitter off at all times. If the personnel are at a remote location, the control system must provide this capability continuously or must include an alternate method of acquiring control that can satisfy the requirement of paragraph (d) of this section that operation be terminated within 3 minutes.

(c) The licensee must establish monitoring procedures and schedules for the station and the indicating instruments employed must comply with Sec. 73.1215.

(1) Monitoring procedures and schedules must enable the licensee to determine compliance with Sec. 73.1560 regarding operating power and AM station mode of operation, Sec. 73.1570 regarding modulation

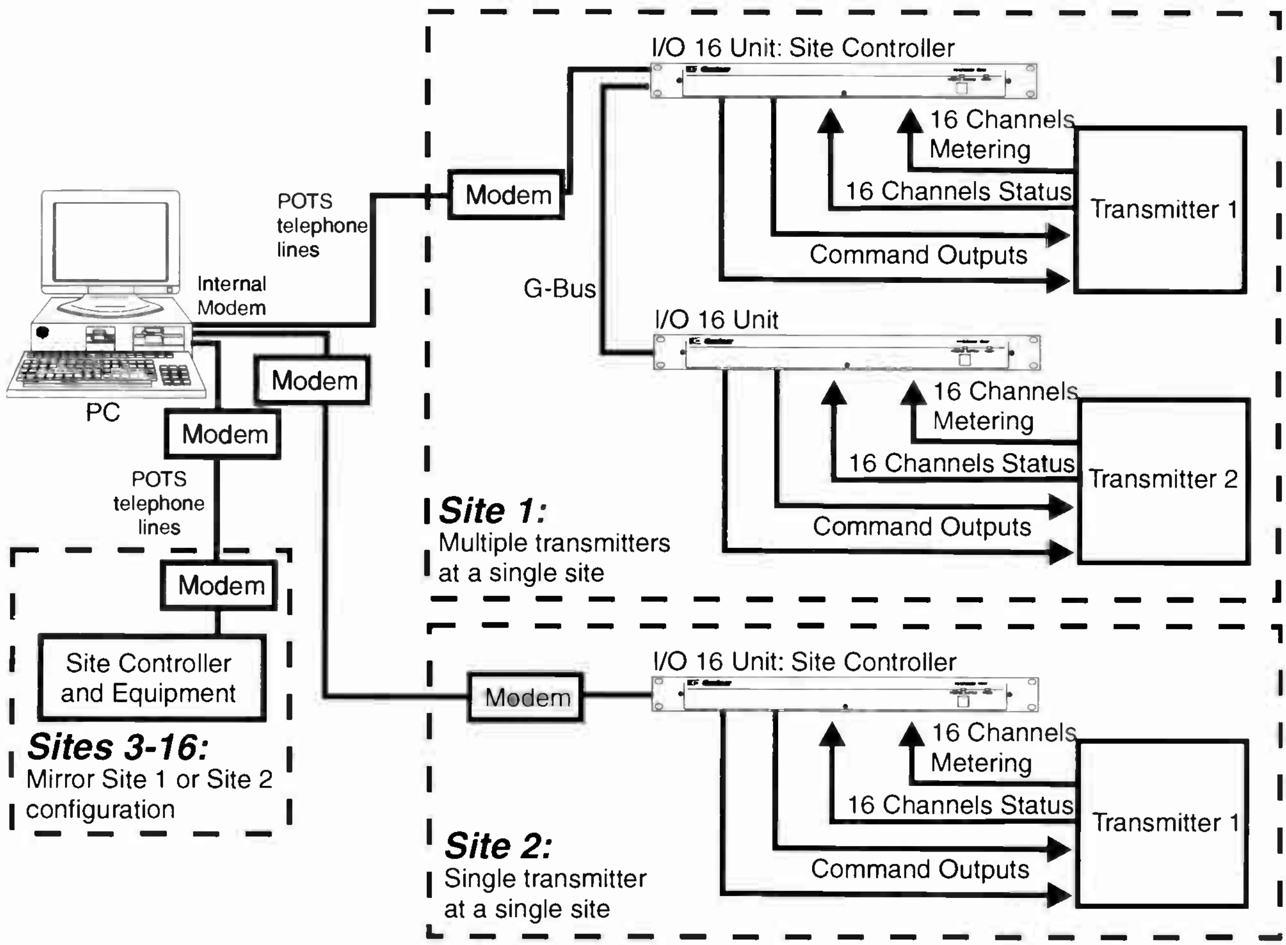


Figure 6.5-9. Overview of networking-type transmitter remote control equipment at 16 sites using POTS interconnection.

levels, and, where applicable, Sec. 73.1213 regarding antenna tower lighting, and Sec. 73.69 regarding the parameters of an AM directional antenna system.

(2) Monitoring equipment must be periodically calibrated so as to provide reliable indications of transmitter operating parameters with a known degree of accuracy. Errors inherent in monitoring equipment and the calibration procedure must be taken into account when adjusting operating parameters to ensure that the limits imposed by the technical rules and the station authorization are not exceeded.

(d) In the event that a broadcast station is operating in a manner that is not in compliance with the technical rules set forth elsewhere in this part or the terms of the station authorization, and the condition is not listed in paragraph (e) of this section, broadcast operation must be terminated within three hours.

(1) Examples of conditions that require termination of operation include excessive power or excessive modulation.

(2) Additional examples for AM stations are any mode of operation not specified by the station license for the pertinent time of day or hours of operation and any condition of antenna parameters or monitoring points out of the tolerances specified elsewhere in this part or by the station's instrument of authorization. For these conditions, operation must be terminated within three minutes unless antenna input power is reduced sufficiently to eliminate any excess radiation.

(3) For AM stations using directional arrays, additional procedures apply when array operating parameters are at variance, monitoring points exceed specified limits, or authorized directional mode capability is lost. See Sec. 73.62, Directional antenna system tolerances; Sec. 73.158, Directional antenna monitoring points; and Sec. 73.1680(b), Emergency antennas.

(e) If a broadcast station is operating in a manner that is not in compliance with one of the following technical rules, operation may continue if the station complies with relevant alternative provisions in the specified rule section.

(1) AM directional antenna system tolerances, see Sec. 73.62;

(2) AM directional antenna monitoring points, see Sec. 73.158;

(3) TV visual waveform, see Sec. 73.691(b);

(4) Reduced power operation, see Sec. 73.1560(d);
(5) Reduced modulation level, see Sec. 73.1570(a);
(6) Emergency antennas, see Sec. 73.1680.

(f) The transmission system must be maintained and inspected in accordance with Sec. 73.1580.

(g) Whenever a transmission system control point is established at a location other than at the main studio or transmitter, notification of that location must be sent to the FCC in Washington, D.C. within 3 days of the initial use of that point. This notification is not required if responsible station personnel can be contacted at the transmitter or studio site during hours of operation.

(h) The licensee must ensure that the station is operated in compliance with Part 11 of this chapter, the rules governing the Emergency Alert System (EAS).

§ 73.1400 Transmission System Monitoring and Control.

The licensee of an AM, FM or TV station is responsible for assuring that at all times the station operates within tolerances specified by applicable technical rules contained in this part and in accordance with the terms of the station authorization. Any method of complying with applicable tolerances is permissible. The following are typical methods of transmission system operation:

(a) Attended operation:

(1) Attended operation consists of ongoing supervision of the transmission facilities by a station employee or other person designated by the licensee. Such supervision may be accomplished by either:

(i) Direct supervision and control of transmission system parameters by a person at the transmitter site; or

(ii) Remote control of the transmission system by a person at the main studio or other location. The remote control system must provide sufficient transmission system monitoring and control capability so as to ensure compliance with Sec. 73.1350.

(2) A station may also be monitored and controlled by an automatic transmission system (ATS) that is configured to contact a person designated by the licensee in the event of a technical malfunction. An automatic transmission system consists of monitoring devices, control and alarm circuitry, arranged so that they interact automatically to operate the station's transmitter and maintain technical parameters within licensed values.

(3) A hybrid system containing some remote control and some ATS features is also permissible.

(4) In the case of remote control or ATS operation, not every station parameter need be monitored or controlled if the licensee has good reason to believe that its stability is so great that its monitoring and control are unnecessary.

(b) Unattended operation: Unattended operation is either the absence of human supervision or the substitution of automated supervision of a station's transmission system for human supervision. In the former case, equipment is employed which is expected to operate within assigned tolerances for extended periods of time. The latter consists of the use of a self-monitoring or ATS-monitored and controlled transmission system that, in lieu of contacting a person designated by the licensee, automatically takes the station off the air within three hours of any technical malfunction which is capable of causing interference.

§ Sec. 73.1580 Transmission System Inspections.

Each AM, FM, and TV station licensee or permittee must conduct periodic complete inspections of the transmitting system and all required monitors to ensure proper station operation.

§ Sec. 73.1635 Special Temporary Authorizations (STA).

(a) A special temporary authorization (STA) is the authority granted to a permittee or licensee to permit the operation of a broadcast facility for a limited period at a specified variance from the terms of the station authorization or requirements of the FCC rules applicable to the particular class of station.

(1) A request for a STA should be filed with FCC in Washington, DC at least 10 days prior to the date of the proposed operation.

(2) The request is to be made by letter and shall fully describe the proposed operation and the necessity for the requested STA. Such letter requests shall be signed by the licensee or the licensee's representative.

(3) A request for a STA necessitated by unforeseen equipment damage or failure may be made without regard to the procedural requirements of this section (e.g. via telegram or telephone). Any request made pursuant to this paragraph shall be followed by a written confirmation request conforming to the requirements of paragraph (a)(2) of this section. Confirmation requests shall be submitted within 24 hours. (See also Sec. 73.1680 Emergency Antennas).

(4) An STA may be granted for an initial period not to exceed 180 days. A limited number of extensions of such authorizations may be granted for additional periods not exceeding 180 days per extension. An STA necessitated by technical or equipment problems, however, may, in practice, be granted for an initial period not to exceed 90 days with a limited number of extensions not to exceed 90 days per extension. The permittee or licensee must demonstrate that any further extensions requested are necessary and that all steps to resume normal operation are being undertaken in an expeditions and timely fashion. The license of a broadcasting station that fails to transmit broadcast signals for any consecutive 12-month period expires as a matter of law at the end of that period, notwithstanding any STA or provision, term, or condition of the license to the contrary.

(5) Certain rules specify special considerations and procedures in situations requiring an STA or permit temporary operation at variance without prior authorization from the FCC when notification is filed as prescribed in the particular rules. See Sec. 73.62, Directional antenna system tolerances; Sec. 73.157, Antenna testing during daytime; Sec. 73.158, Directional

antenna monitoring points; Sec. 73.691, Visual modulation monitoring; Sec. 73.1250, Broadcasting emergency information; Sec. 73.1350, Transmission system operation; Sec. 73.1560, Operating power and mode tolerances; Sec. 73.1570, Modulation levels: AM, FM, and TV aural; Sec. 73.1615, Operation during modification of facilities; Sec. 73.1680, Emergency antennas; and Sec. 73.1740, Minimum operating schedule.

(b) An STA may be modified or cancelled by the FCC without prior notice or right to hearing.

(c) No request by an AM station for temporary authority to extend its hours of operation beyond those authorized by its regular authorization will be accepted or granted by the FCC except in emergency situations conforming with the requirements of Sec. 73.3542, Application for Emergency Authorization. See also Sec. 73.1250, Broadcasting Emergency Information.

§ Sec. 73.1820 Station Log.

(a) Entries must be made in the station log either manually by a person designated by the licensee who is in actual charge of the transmitting apparatus, or by automatic devices meeting the requirements of paragraph (b) of this section. Indications of operating parameters that are required to be logged must be logged prior to any adjustment of the equipment. Where adjustments are made to restore parameters to their proper operating values, the corrected indications must be logged and accompanied, if any parameter deviation was beyond a prescribed tolerance, by a notation describing the nature of the corrective action. Indications of all parameters whose values are affected by the modulation of the carrier must be read without modulation. The actual time of observation must be included in each log entry. The following information must be entered:

(1) All stations: (i) Entries required by Sec. 17.49 of this chapter concerning any observed or otherwise known extinguishment or improper functioning of a tower light:

(A) The nature of such extinguishment or improper functioning.

(B) The date and time the extinguishment or improper operation was observed or otherwise noted.

(C) The date, time and nature of adjustments, repairs or replacements made.

(ii) Any entries not specifically required in this section, but required by the instrument of authorization or elsewhere in this part.

(iii) An entry of each test and activation of the Emergency Alert System (EAS) pursuant to the requirement of part 11 of this chapter and the EAS Operating Handbook. Stations may keep EAS data in a special EAS log which shall be maintained at a convenient location; however, this log is considered a part of the station log.

(2) Directional AM stations without an FCC-approved antenna sampling system (See Sec. 73.68):

(i) An entry at the beginning of operations in each mode of operation, and thereafter at intervals not exceeding 3 hours, of the following (actual readings observed prior to making any adjustments to the equipment and an indication of any corrections to restore parameters to normal operating values):

(A) Common point current.

(B) When the operating power is determined by the indirect method, the efficiency factor F and either the product of the final amplifier input voltage and current or the calculated antenna input power. See Sec. 73.51(e).

(C) Antenna monitor phase or phase deviation indications.

(D) Antenna monitor sample currents, current ratios, or ratio deviation indications.

(ii) Entries required by Sec. 73.61 performed in accordance with the schedule specified therein.

(iii) Entries of the results of calibration of automatic logging devices (see paragraph (b) of this section) or indicating instruments (see Sec. 73.67), whenever performed.

(b) Automatic devices accurately calibrated and with appropriate time, date and circuit functions may be utilized to record entries in the station log Provided:

(1) The recording devices do not affect the operation of circuits or accuracy of indicating instruments of the equipment being recorded;

(2) The recording devices have an accuracy equivalent to the accuracy of the indicating instruments;

(3) The calibration is checked against the original indicators as often as necessary to ensure recording accuracy;

(4) In the event of failure or malfunctioning of the automatic equipment, the person designated by the licensee as being responsible for the log small make the required entries in the log manually at that time;

(5) The indicating equipment conforms to the requirements of Sec. 73.1215 (Indicating instruments—specifications) except that the scales need not exceed 5 cm (2 inches) in length. Arbitrary scales may not be used.

(c) In preparing the station log, original data may be recorded in rough form and later transcribed into the log.

§ Sec. 73.1870 Chief Operators.

(a) The licensee of each AM, FM, or TV broadcast station must designate a person to serve as the station's chief operator. At times when the chief operator is unavailable or unable to act (e.g., vacations, sickness), the licensee shall designate another person as the acting chief operator on a temporary basis.

(b) Chief operators shall be employed or serve on the following basis:

(1) The chief operator for an AM station using a directional antenna or operating with greater than 10 kW authorized power, or of a TV station is to be an employee of the station on duty for whatever number of hours each week the station licensee determines is necessary to keep the station's technical operation in compliance with FCC rules and the terms of the station authorization.

(2) Chief operators for non-directional AM stations operating with authorized powers not exceeding 10 kW and FM stations may be either an employee of the station or engaged to serve on a contract basis for whatever number of hours each week the licensee determines is necessary to keep the station's technical operation in compliance with the FCC rules and terms of the station authorization.

(3) The designation of the chief operator must be in writing with a copy of the designation posted with the station license. Agreements with chief operators serving on a contract basis must be in writing with a copy kept in the station files.

(c) The chief operator is responsible for completion of the following duties specified in this paragraph below. When these duties are delegated to other persons, the chief operator shall maintain supervisory oversight sufficient to know that each requirement has been fulfilled in a timely and correct manner.

(1) Inspections and calibrations of the transmission system, required monitors, metering and control systems; and any necessary repairs or adjustments where indicated. (See Sec. 73.1580.)

(2) Periodic AM field monitoring point measurements, equipment performance measurements, or other tests as specified in the rules or terms of the station license.

(3) Review of the station records at least once each week to determine if required entries are being made correctly. Additionally, verification must be made that the station has been operated as required by the rules or the station authorization. Upon completion of the review, the chief operator or his designee must date and sign the log, initiate any corrective action which may be necessary, and advise the station licensee of any condition which is repetitive.

(4) Any entries which may be required in the station records. (See Sec. 73.1820.)

6.6
WAVEGUIDE

JERRY WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

Any structure that guides electromagnetic waves can be considered a *waveguide*. Most often, however, this term refers to closed metal cylinders that maintain the same cross-sectional dimensions over long distances.

As the operating frequency of a system reaches into the UHF band, waveguide based transmission line systems become practical. From the mechanical standpoint, waveguide is simplicity itself. There is no inner conductor; RF energy is *launched* into the structure and propagates to the load. Several types of waveguide are available, including rectangular, square, circular and elliptical. Waveguide offers several advantages over coax. First, unlike coax, waveguide can carry more power as the operating frequency increases. Second, efficiency is significantly greater with waveguide at higher frequencies.

Rectangular waveguide is commonly used in high-power transmission systems. Circular waveguide may also be used, especially for applications requiring a cylindrical member, such as a rotating joint for an antenna feed. The physical dimensions of the guide are selected to provide for propagation in the dominant (lowest-order) mode.

Waveguide is not without its drawbacks, however. Rectangular or square guide constitutes a large windload surface, which places significant structural demands on a tower. Because of the physical configuration of rectangular and square guide, pressurization is limited, depending on the type of waveguide used (0.5 psi is typical). Excessive pressure can deform the guide shape and result in increased VSWR. Wind also may cause deformation and ensuing VSWR problems. These considerations have led to the development of circular and elliptical waveguide.

PROPAGATION MODES

Propagation modes for waveguide fall into two broad categories:

- Transverse electric (TE) waves
- Transverse magnetic (TM) waves

With TE waves, the electric vector (*E vector*) is perpendicular to the direction of propagation. With TM waves, the magnetic vector (*H vector*) is perpendicular to the direction of propagation. These propagation modes take on integers (from 0 or 1 to ∞) that define field configurations. Only a limited number of these modes can be propagated, depending on the dimensions of the guide and the operating frequency.

Energy cannot propagate in waveguide unless the operating frequency is above the *cutoff frequency*. The cutoff frequency for rectangular guide is:

$$F_c = \frac{C}{2 \times a}$$

Where:
F_c = waveguide cutoff frequency
C = 1.179×10^{10} in/s (the velocity of light)
a = the wide dimension of the guide

The cutoff frequency for circular waveguide is defined by the following equation:

$$F_c = \frac{C}{3.14 \times a'}$$

Where: a' = the radius of the guide

There are four common propagation modes in waveguide:

- $TE_{0,1}$, the principle mode in rectangular waveguide
- $TE_{1,0}$, also used in rectangular waveguide
- $TE_{1,1}$, the principle mode in circular waveguide. $TE_{1,1}$ develops a complex propagation pattern with electric vectors curving inside the guide. This mode exhibits the lowest cutoff frequency of all modes, which allows a smaller guide diameter for a specified operating frequency
- $TM_{0,1}$, which has a slightly higher cutoff frequency than $TE_{1,1}$ for the same size guide. Developed as a result of discontinuities in the waveguide, such as flanges and transitions, $TM_{0,1}$ energy is not coupled out by either dominant or cross-polar transitions. The parasitic energy must be filtered out, or the waveguide diameter picked carefully to reduce the unwanted mode.

The field configuration for the dominant mode in rectangular waveguide is illustrated in Figure 6.6-1. Note that the electric field is vertical, with intensity maximum at the center of the guide and dropping off sinusoidally to zero intensity at the edges. The magnetic field is in the form of loops, which lie in planes that are at right angles to the electric field (parallel to the top and bottom of the guide). The magnetic field

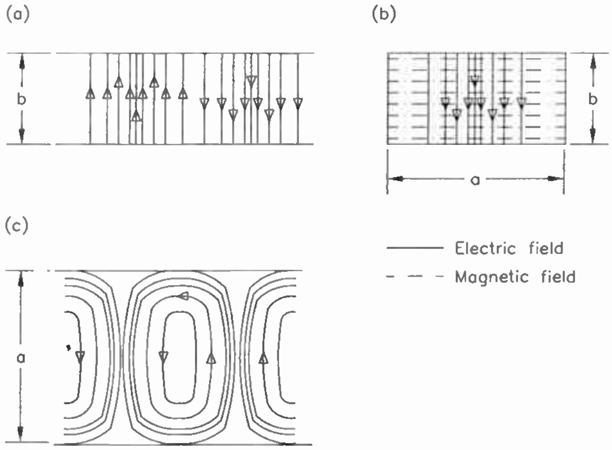


Figure 6.6-1. Field configuration of the dominant or $TE_{1,0}$ mode in a rectangular waveguide: (a) X axis, (b) Y axis, (c) Z axis.

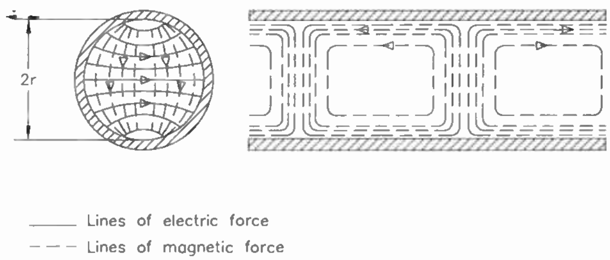


Figure 6.6-2. Field configuration of the dominant mode in circular waveguide.

distribution is the same for all planes perpendicular to the *Y*-axis. In the *X* direction, the intensity of the component of magnetic field that is transverse to the axis of the waveguide (the component in the direction of *X*) is at any point in the waveguide directly proportional to the intensity of the electric field at that point. This entire configuration of fields travels in the direction of the waveguide axis (the *Z* direction in Figure 6.6-1).

The field configuration for the $TE_{1,1}$ mode in circular waveguide is illustrated in Figure 6.6-2. The $TE_{1,1}$ mode has the longest cutoff wavelength and is, accordingly, the dominant mode. The next higher mode is the $TM_{0,1}$, followed by $TE_{2,1}$.

Dual Polarity Waveguide

Waveguide will support dual polarity transmission within a single run of line. A combining element (*dual polarized transition*) is used at the beginning of the run, and a splitter (polarized transition) is used at the end of the line. Square waveguide has found numerous applications in such systems. The $TE_{1,0}$ and $TE_{0,1}$ modes are theoretically capable of propagation without cross coupling, at the same frequency, in lossless waveguide of square cross section. In practice, surface irregularities, manufacturing tolerances and wall losses give rise to $TE_{1,0}$ and $TE_{0,1}$ mode cross-conversion. Because this conversion occurs continually along the waveguide, long guide runs are usually avoided in dual polarity systems.

Efficiency

Waveguide loses result from the following:

- Power dissipation in the waveguide walls and the dielectric material filling the enclosed space
- Leakage through the walls and transition connections of the guide
- Localized power absorption and heating at the connection points.

The operating power of waveguide may be increased through pressurization. Sulfur hexafluoride is commonly used as the pressurizing gas.

Ridged Waveguide

Rectangular waveguide may be ridged to provide a lower cutoff frequency, and thereby permit use over a wider frequency band. One and two-ridged guides are used (see Figure 6.6-3). Increased bandwidth comes at the expense of increased attenuation, relative to an equivalent section of rectangular guide.

Circular Waveguide

Circular waveguide offers several mechanical benefits over rectangular or square guide. The wind load of circular guide is 2/3 that of rectangular waveguide. It also presents lower and more uniform windloading

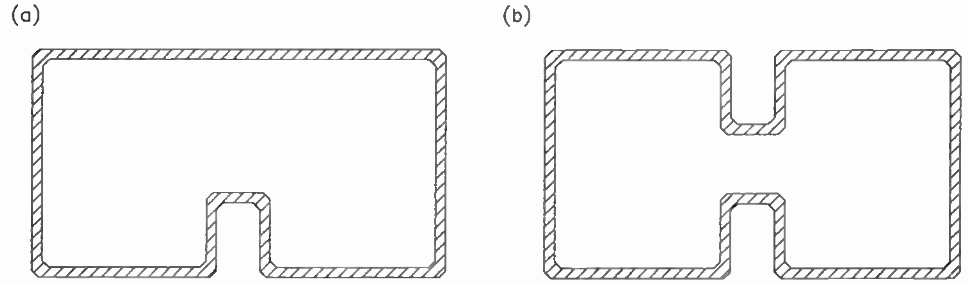


Figure 6.6-3. Ridged waveguide: (*a*) single-ridged; (*b*) double-ridged.

than rectangular waveguide, reducing tower structural requirements.

The same physical properties of circular waveguide that give it good power handling and low attenuation also result in electrical complexities. Circular waveguide has two potentially unwanted modes of propagation, the cross-polarized $TE_{1,1}$ and $TM_{0,1}$ modes.

Circular waveguide, by definition, has no short or long dimension and, consequently, no method to prevent the development of *cross-polar* or *orthogonal* energy. Cross-polar energy is formed by small ellipticities in the waveguide. If the cross-polar energy is not trapped out, the parasitic energy can recombine with the dominant mode energy.

Parasitic Energy

Hollow circular waveguide works as a high-Q resonant cavity for some energy and as a transmission medium for the rest. The parasitic energy present in the cavity formed by the guide will appear as increased VSWR if not disposed of. The polarization in the guide meanders and rotates as it propagates from the source to the load. The end pieces of the guide, typically circular-to-rectangular transitions, are polarization sensitive (see Figure 6.6-4(a)). If the polarization of the incidental energy is not matched to the transition, energy will be reflected.

Several factors can result in this undesirable polarization. One cause is out-of-round guides that result from non-standard manufacturing tolerances. In Figure 6.6-4, the solid lines depict the situation at launching: perfectly circular guide with perpendicular polarization. The dashed lines show how certain ellipticities cause polarization rotation into unwanted states, while others have no effect. A 0.2% change in diameter can produce a −40 dB cross-polarization component per wavelength. This is roughly 0.03 in. for 18 in. of guide length.

Other sources of cross-polarization include twisted and bent guides, out-of-roundness, offset flanges, and transitions. Various methods are used to dispose of this energy trapped in the cavity, including absorbing loads placed at the ground and/or antenna level.

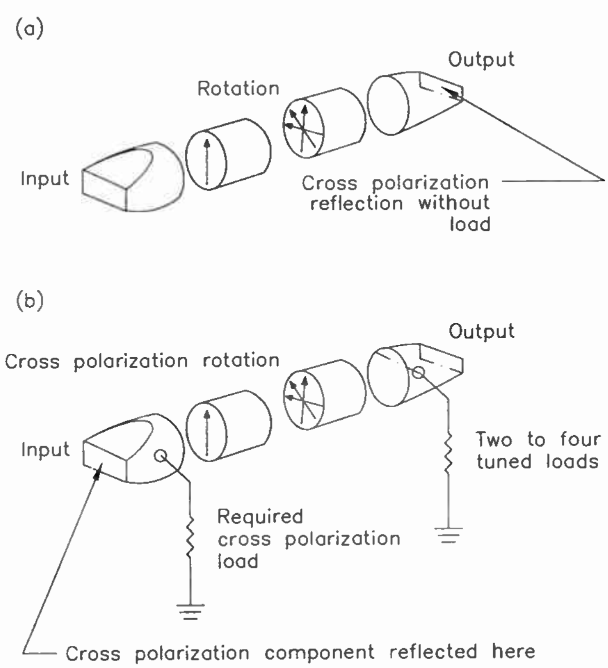


Figure 6.6-4. The effects of parasitic energy in circular waveguide: (*a*) trapped cross-polarization energy; (*b*) delayed transmission of the trapped energy.

Doubly-Truncated Waveguide

The design of *doubly truncated waveguide* (DTW) is intended to overcome the problems that may result from parasitic energy in a circular waveguide. As shown in Figure 6.6-5, DTW consists of an almost elliptical guide inside a circular shell. This guide does not support cross-polarization; tuners and absorbing loads are not required. The low windload of hollow circular guide is maintained, except for the flange area.

Each length of waveguide is actually two separate pieces: a doubly truncated center section and a circular outer skin, joined at the flanges on each end. A large hole in the broadwall serves to pressurize the circular outer skin. Equal pressure inside the DTW and inside the circular skin ensures that the guide will not breathe or buckle as a result of rapid temperature changes.

DTW exhibits about 3% higher windloading than circular waveguide (because of the transition section at the flange joints), and 32% lower loading than comparable rectangular waveguide.

Impedance Matching

The efficient flow of power from one type of transmission medium to another requires matching of the field patterns across the boundary to launch the wave into the second medium with a minimum of reflections. Coaxial line is typically matched into rectangular

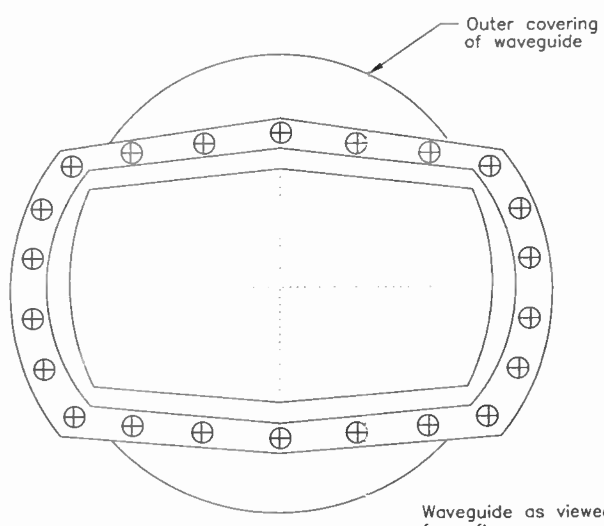


Figure 6.6-5. Physical construction of doubly-truncated waveguide.

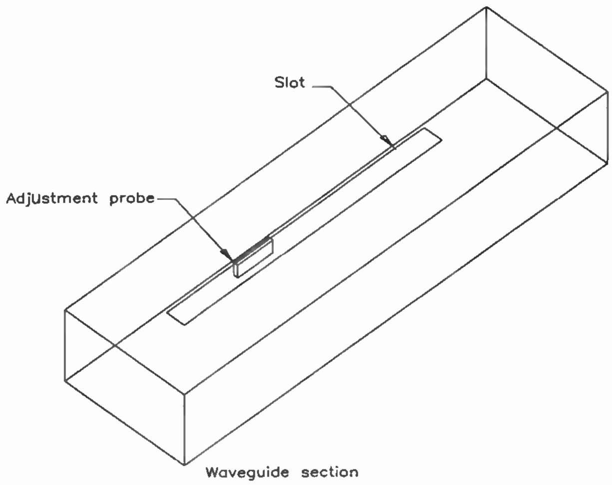


Figure 6.6-6. A probe configured to introduce a reflection in a waveguide that is adjustable in magnitude and phase.

waveguide by extending the center conductor of the coax through the broadwall of the guide, parallel to the electric field lines across the guide. Alternatively, the center conductor can be formed into a loop and oriented to couple the magnetic field to the guide mode.

Standing waves are generally to be avoided in waveguide for the same reasons that they are to be avoided in coaxial transmission lines. Accordingly, it is usually necessary to provide impedance matching systems in waveguides to eliminate standing waves. One approach involves the introduction of a compensating reflection in the vicinity of a load which neutralizes the standing waves that would exist in the system because of an imperfect match. A probe or tuning screw is commonly used to accomplish this, as illustrated in Figure 6.6-6. The tuning screw projects into the waveguide in a direction parallel to the electric field. This is equivalent to shunting a capacitive load across the guide. The susceptance of the load increases with extension into the guide up to ¼ wavelength. When the probe is exactly ¼ wavelength long, it becomes resonant and causes the guide to behave as though there was an open circuit at the point of the resonant probe. Probes longer than ¼ wavelength but shorter than ¾ wavelength introduce inductive loading. The extent to which such a probe projects into the waveguide determines the compensating reflection, while the position of the probe with respect to the standing wave pattern to be eliminated determines the phasing of the reflected wave.

Dielectric slugs produce an effect similar to that of a probe. The magnitude of the effect depends upon the following considerations:

- The dielectric constant of the slug
- Thickness of the slug in an axial direction
- Whether the slug extends entirely across the waveguide.

The phase of the reflected wave is controlled by varying the axial position of the slug.

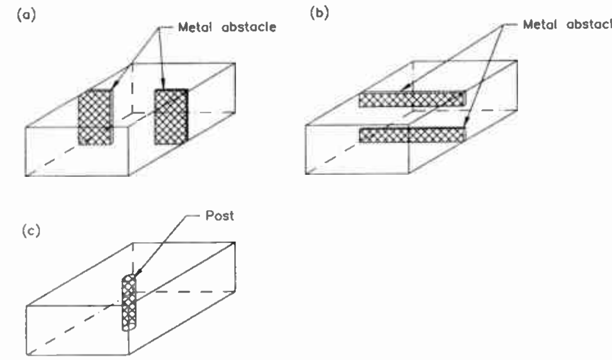


Figure 6.6-7. Waveguide obstructions used to introduce compensating reflections: (*a*) inductive window, (*b*) capacitive window, (*c*) post (inductive) element.

There are several alternatives to the probe and slug for introducing controllable irregularity for impedance matching, including a metallic barrier or *window* placed at right angles to the axis of the guide, (see Figure 6.6-7). Three configurations are shown:

- The arrangement illustrated in Figure 6.6-7(a) produces an effect equivalent to shunting the waveguide with an inductive reactance
- The arrangement shown in Figure 6.6-7(b) produces the effect of a shunt capacitive susceptance
- The arrangement shown in Figure 6.6-7(c) produces an inductive shunt susceptance.

The waveguide equivalent of the coaxial cable tuning stub is a *T* section, (see Figure 6.6-8). The magnitude of the compensating effect is controlled by the position of the short-circuiting plug in the branch. The phase of the compensating reflected wave produced by the branch is determined by the position of the branch in the guide.

Waveguide Filters

A section of waveguide beyond cutoff constitutes a simple high pass reflective filter. Loading elements in the form of posts or stubs may be employed to supply the reactances required for conventional lumped-constant-filter designs.

Absorption filters avoid the reflection of unwanted

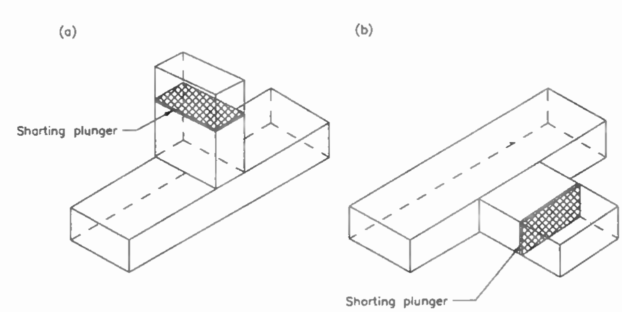


Figure 6.6-8. Waveguide stub elements used to introduce compensating reflections: (*a*) series *T* element, (*b*) shunt *T* element.

energy by incorporating lossy material in secondary guides that are coupled through so called *leaky walls* (small sections of guide beyond cutoff in the passband). Such filters are typically used to suppress harmonic energy.

SYSTEM CONSIDERATIONS

Waveguide system installation is both easier and more difficult than traditional transmission line. There is no inner conductor to align, but alignment pins and more bolts are required per flange. Transition hardware to accommodate loads and coaxial-to-waveguide interfacing is also required.

Flange reflections can add up in phase at certain frequencies, resulting in high VSWR. The length of the guide must be chosen so that flange reflection buildup does not occur within the operating bandwidth.

Flexible sections of waveguide are used to join rigid sections or components that cannot be otherwise aligned. Flexible sections also permit controlled physical movement resulting from thermal expansion of the line. Such hardware is available in a variety of forms. Corrugated guide is commonly produced by shaping thin-wall seamless rectangular tubing. Flexible waveguide can accommodate only a limited amount of mechanical movement. Depending on the type of link, the manufacturer may specify a maximum number of bends.

Tuning

Circular waveguide must be tuned. This process is a two-step procedure. First, the cross-polar $TE_{1,1}$ component is reduced, primarily through *axial ratio compensators* or *mode optimizers.* These devices counteract the net system ellipticity and indirectly minimize cross-polar energy. The cross-polar filters may also be rotated to achieve maximum isolation between the dominant and cross-polar modes. Cross-polar energy manifests itself as a net signal rotation at the end of the waveguide run. A perfect system would have a net rotation of zero.

In the second step, tuning slugs at both the top and bottom of the waveguide run are adjusted to reduce the overall system VSWR. Tuning waveguide can be a complicated and time consuming procedure, but—normally—once set, tuning does not drift and must be redone only if major component changes are made.

Waveguide Hardware

Increased use of waveguide has led to the development of waveguide based hardware for all elements from the output of the transmitter to the load. Waveguide based filters, elbows, directional couplers, switches, combiners and diplexers are available. The RF performance of a waveguide component is usually better than the same item in coax. This is especially true in the case of diplexers and filters. Waveguide based hardware provides lower attenuation and greater power handling capability for a given physical size.

Cavity Resonators

Any space completely enclosed by conducting walls can contain oscillating electromagnetic fields. Such a cavity possesses certain frequencies at which it will resonate when excited by electrical oscillations. These *cavity resonators* find extensive use as resonant circuits at VHF and above. Advantages of cavity resonators over conventional LC circuits include:

- Simplicity in design
- Relatively large physical size compared with alternative methods of obtaining resonance. This attribute is important in high-power applications
- High Q
- The cavity can be configured to develop an extremely high shunt impedance.

Cavity resonators are commonly used at wavelengths on the order of 10 cm or less.

The simplest cavity resonator is a section of waveguide shorted at each end with a length l equal to

$$l = \frac{\lambda_g}{2}$$

Where λ_g = the guide wavelength

This configuration results in a resonance similar to that of a ½ wavelength transmission line short-circuited at the receiving end. A sphere or any other enclosed surface (irrespective of how irregular the outline) may also be used to form a cavity resonator.

Any given cavity is resonant at a number of frequencies, corresponding to the different possible field conditions that can exist within the space. The resonance having the longest wavelength (lowest frequency) is termed the *dominant* or *fundamental resonance.* The resonant wavelength is proportional to the size of the resonator. If all dimensions are doubled, the wavelength corresponding to resonance will likewise be doubled. The resonant frequency of a cavity can be changed by incorporating one or more of the following mechanisms:

- Altering the mechanical dimensions of the cavity. Small changes can be achieved by flexing walls, while large changes require some form of sliding member
- Coupling reactance into the resonator through a coupling loop
- Introducing a movable copper paddle into the cavity. A paddle placed inside the resonator will affect the normal distribution of flux and tend to raise the resonant frequency by an amount determined by the orientation of the paddle.

The Q of a cavity resonator has the same significance as for a conventional resonant circuit. Q can be defined for a cavity by the following relationship:

$$Q = 2\pi \frac{E_s}{E_l}$$

Where E_s = energy stored and E_l = energy lost per cycle.

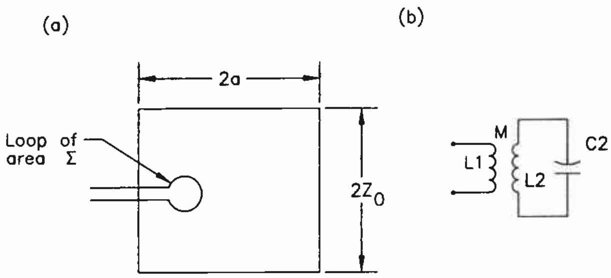


Figure 6.6-9. Cavity resonator coupling: (*a*) coupling loop, (*b*) equivalent circuit.

The energy stored is proportional to the square of the magnetic flux density integrated throughout the volume of the resonator. The energy lost in the walls is proportional to the square of the magnetic flux density integrated over the surface of the cavity. In order to obtain high Q, the resonator should have a large ratio of volume to surface area, because it is the volume that stores energy and the surface area that dissipates energy.

Coupling can be obtained from a resonator by means of a coupling loop or coupling electrode. Magnetic coupling is accomplished through the use of a loop oriented so as to enclose magnetic flux lines existing in the desired mode of operation. This technique is illustrated in Figure 6.6-9. A current passed through the loop will excite oscillations of this mode. Conversely, oscillations existing in the resonator will induce a voltage in the coupling loop. The magnitude of the coupling can be controlled by rotating the loop; the coupling reduces to zero when the plane of the loop is parallel to the magnetic flux. Coupling of a resonator may also be accomplished through the use of a probe or opening in one wall of the cavity.

SYSTEM LAYOUT CONSIDERATIONS[1]

Typical system configurations for different types of waveguide are shown in Figure 6.6-10.

Because of the large size of bends, transitions, etc., and the differential expansion between waveguide and tower, careful coordination with the tower engineer and waveguide manufacturer is essential to prevent problems during installation and operation.

In addition, the horizontal run must be long enough to withstand the flexing caused by expansion in the much longer vertical run. Figure 6.6-11 provides a guide to determine this length for circular waveguide systems. As can be seen, the horizontal run needs to be about 10% of the vertical for shorter systems and 5% for long systems. Always consult the manufacturer for confirmation prior to finalizing locations of the tower and transmitter building.

[1] This section is an excerpt from Chapter 2.3 "Transmission Lines" from the *NAB Engineering Handbook, 8th edition.* 1992, pp. 153–160.

Horizontal and Vertical Straight Sections

Straight sections range from 11.5 to 12 ft depending on waveguide type and channel of operation. Follow the manufacturer's catalog recommendations.

The most common starting point for developing the list of required straight sections is the waveguide elbow located at the base of the tower. Working toward the building from the bend, the standard lengths recommended by the manufacturer are used back through the building opening. If the dimension of the final mating pieces cannot be determined accurately, specify a *length to be advised* piece. This will alert the manufacturer that a section must be made available based on last minute measurements.

The vertical run is laid out by stacking lengths end-to-end until the appropriate height is reached for positioning the top anchor plate. The same rule for a *to be advised* length should be applied as in the horizontal run.

Additional pieces will be needed between this point and the input to the antenna. As indicated in the figures, these will be a combination of waveguide-to-coax transitions, waveguide bends and special waveguide lengths. Since the waveguide is fastened rigidly at the top of the run, and the antenna is fastened rigidly at its mounting plate, the waveguide in-between must also be allowed to grow and shrink due to thermal expansion and contraction. U-link bends or 90° elbows are the most common method for allowing the necessary *flexing* movement.

Tower Top

The vertical run is secured at the top end with an anchor plate or milkstool which must be carefully located on the tower to line up with the hangers for the vertical run and with the antenna inter-connecting components. The milkstool hanger concept (Figure 6.6-12(a)) provides adjustment in all directions to allow for minor field variations.

Hangers

Thermal expansion in both vertical and horizontal runs, depicted in Figure 6.6-13, causes both to change positions. At the lower portion of the vertical run, lateral braces (Figure 6.6-12(b)) are used to provide lateral support while allowing this shifting of position to occur. The supporting hangers for the horizontal run must also allow for an increased amount of movement at this area.

A spring hanger (Figure 6.6-12(c)) is required on all other sections of waveguide in both the horizontal and vertical runs to provide the expansion/contraction capability necessary. For spring hanger and lateral brace mounting details, the manufacturer's installation bulletins should be thoroughly reviewed. With careful engineering and attention to all the necessary details, the hangers should attach to both the horizontal bridge and the vertical support structure with a minimum of difficulty.

Keep in mind that the coefficients of thermal expan-

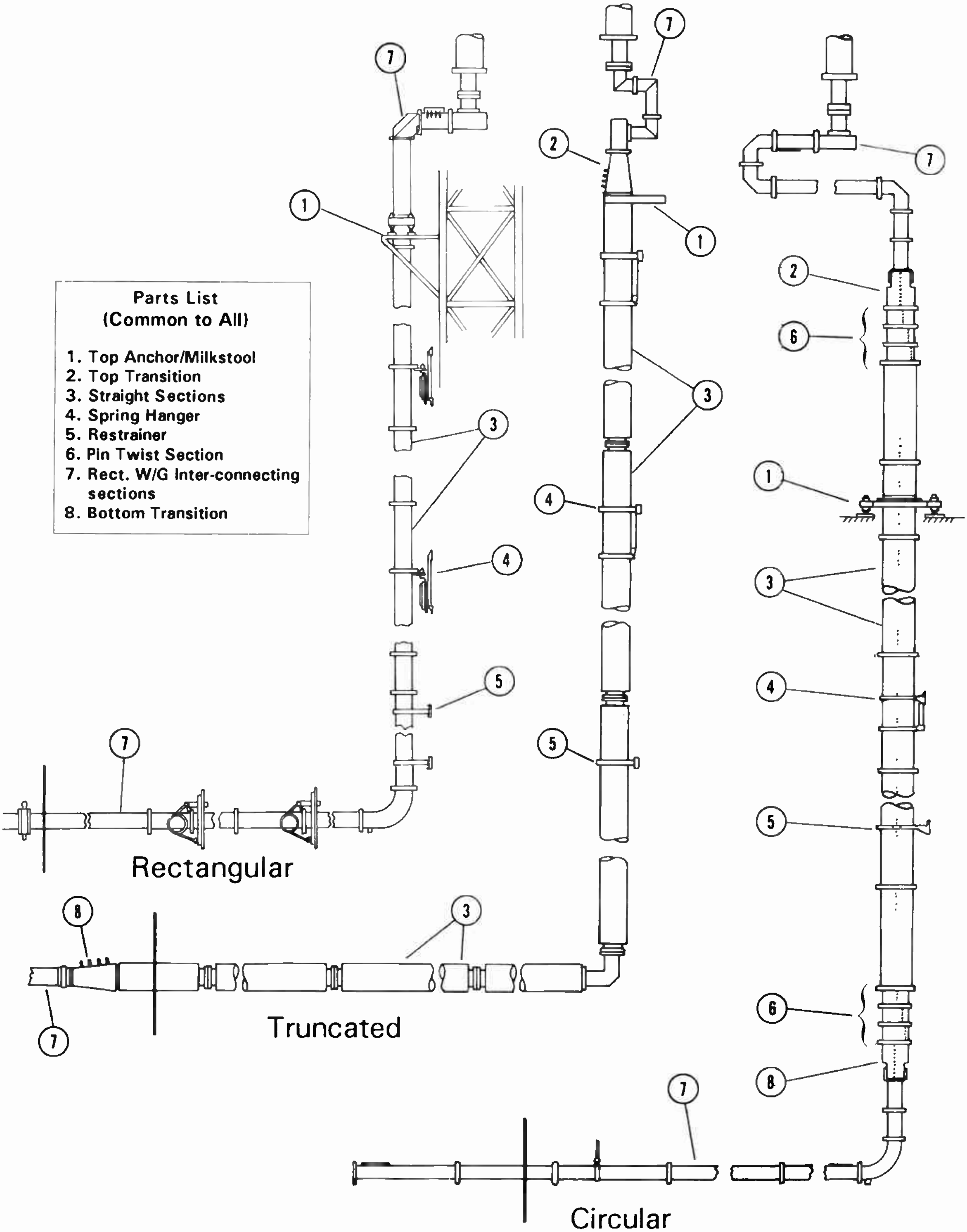


Figure 6.6-10. Typical waveguide system layouts.

Vertical Run Height (feet)	Minimum Horizontal Run Length (feet)					
	GLW			DTW		
	1350	1500	1750	1350	1500	1750
2000	75	90	100	72	84	120
1500	66	77	87	60	84	108
1000	54	61	70	60	72	96
500	38	42	49	48	60	84

Figure 6.6-11. Typical manufacturer's recommendations for determining minimum length of horizontal run to provide for expansion of vertical run.

sion are 13×10^{-6} per degree F for the aluminum waveguide, and approximately 6.5×10^{-6} per degree F for the tower steel. This results in a relative differential movement of about 3/4 in. per 100 ft for a 100°F temperature swing (2.5 cm per 40 m for a 56°C temperature swing). Also, the ambient temperature of the guide and the tower are different due to the additional heating effect of the RF power being transmitted. The resulting expansion/contraction must be accommodated by providing sufficient clearance between all waveguide components and the tower or bridge (see Figure 6.6-13).

Transitions and Associated Components

Rectangular waveguide systems may include transitions as shown in the typical layouts of Figure 6.6-10. Rectangular to coaxial transitions are most commonly used to interface the main waveguide run to the antenna and, if equipped with a coax input, the diplexer. Single polarized circular and doubly truncated waveguide systems will also use transitions from their shape to standard rectangular waveguide at both ends (see Figure 6.6-12(d)).

Single polarized circular systems incorporate multisection pin twists (Figure 6.6-12(e)) at the transitions to facilitate alignment of the electric field (E-field) with the mating rectangular waveguide feeding the antenna and transmitter. This simplifies both the layout of the waveguide system on the tower and the installation. In rectangular only systems, special twist sections are needed to rotate the run as required for proper alignment.

Installation

Pages could be written about the handling and installation practices appropriate for waveguide sections. A summary of the major considerations is presented here.

The crew contracted to do the installation will play a major role in the success of the project. Select one with experience, credentials and recent references. This is not a good place to try to save a few dollars.

If the waveguide will be stored at the site prior to installation, keep the interior surfaces dry and clean by covering the sections with plastic sheeting. Tape the sides and ends securely to keep out wind-driven rain and dirt.

Hanger Installation

Before uncovering any waveguide sections, install all the hangers on the tower and the horizontal bridge. Follow the detail drawings developed by the tower manufacturer to insure the proper interface between the hanger and the tower. Due to the significant windload which the waveguide run will place on the tower, it is important that these instructions be followed carefully.

As shown in the system layout drawings of Figure 6.6-10, several types of hangers will be utilized—a fixed hanger at the top of the tower, spring hangers down the tower, lateral restrainers near the base of the tower and horizontal spring hangers beneath the bridge. Spring hangers with additional travel are provided with certain types of installations to accommodate the larger displacement near the base of the tower in both the vertical and horizontal runs. Be sure that the manufacturer's instructions are followed when installing hangers.

After installation, all hangers must provide the necessary movement to allow waveguide and flange displacement of the type shown in Figure 6.6-13 with no interference at any tower member.

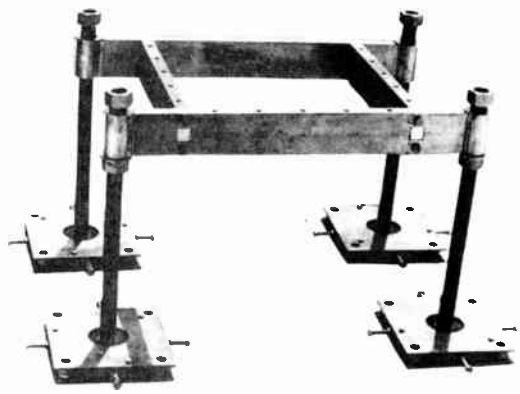
Waveguide Installation

Installation of straight sections usually starts at the base of the tower after properly locating the E-plane bend for clearance at both anticipated temperature extremes. However, single polarized circular waveguide is installed from the top down to eliminate the need for special field-cut circular sections in the vertical run.

Straight sections are normally provided with an O-ring groove in one flange only. This allows the sections to be installed with the groove facing up to provide a cavity for centering the O-ring and prevent pinching when bolting mating flanges together. Check for a uniform gap around the flange perimeter after torquing all flange hardware. A pinched O-ring will ultimately need repair since it will cause VSWR changes, arcing, pressure leaks—or all three.

As mentioned above, be sure the bend at the base of the tower is placed in a location which provides freedom of movement for both the vertical and horizontal runs when differential thermal expansion occurs. As indicated, this differential amount can be approximated using the "3/4 in/100 ft" (2.5 cm/40 m) guideline. Note that this will be true for both the vertical and horizontal runs, assuming steel is used for the tower and the horizontal bridge. This general rule is based on a 100°F (56°C) change in temperature. Thus, if the temperature is expected to increase 50°F (28°C) from the ambient at time of installation, the amount of differential *expansion* will be one-half that previously indicated. However, remember that *shrinkage* will also occur as the temperature drops 50° (28°C) from the ambient, so the same amount of upward movement must also be accommodated.

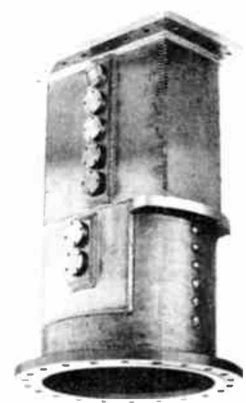
Waveguide flanges in mating sections are normally



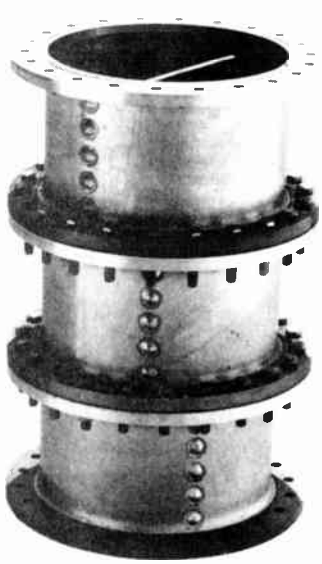
A) "Milkstool" Type Top Anchor

B) Lateral Restrainer for Circular Waveguide

C) Spring Hanger for Circular Waveguide



D) Circular to Rectangular Transition



E) Circular Waveguide Pin Twist Section

Figure 6.6-12. Typical waveguide system components.

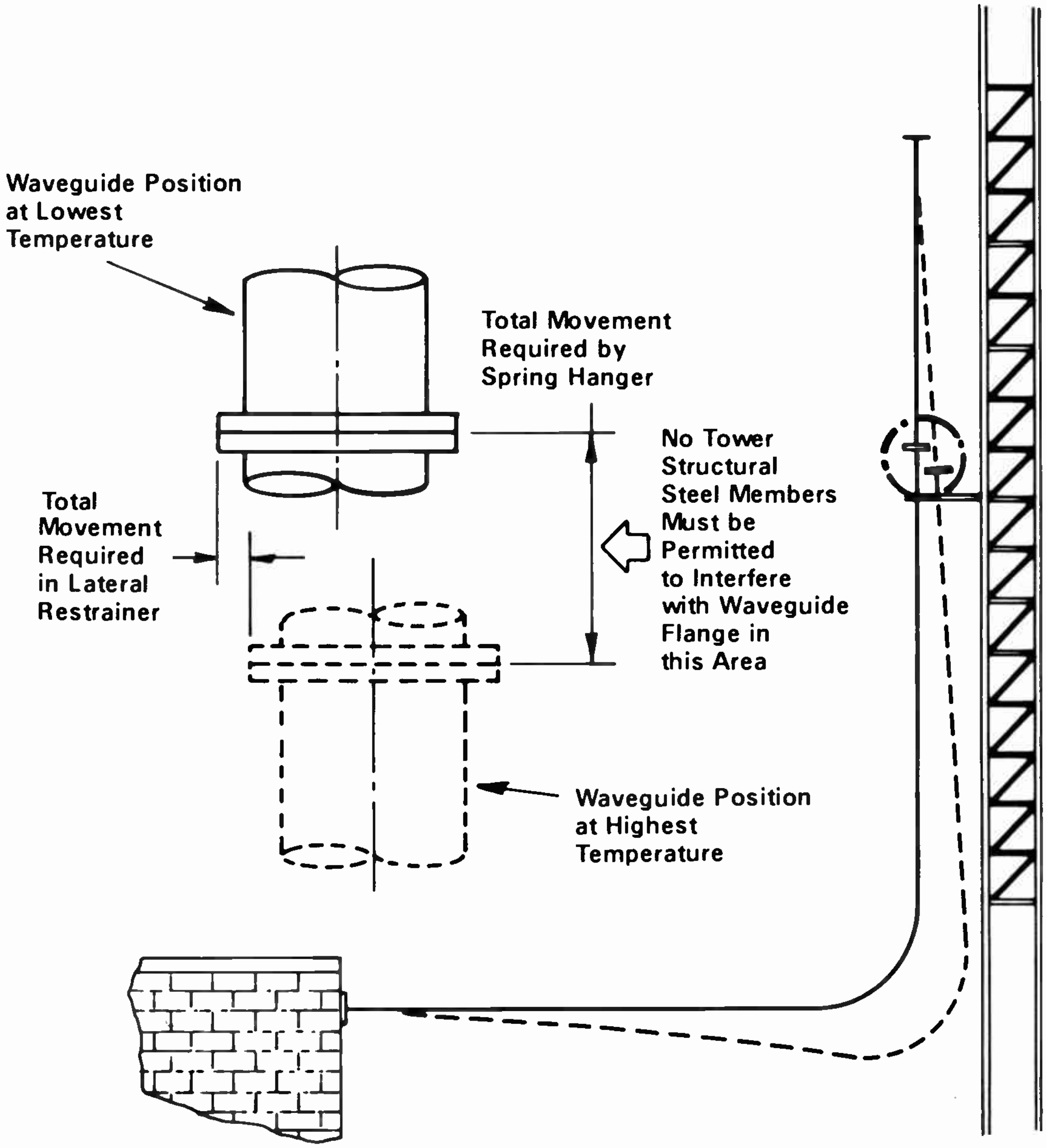


Figure 6.6-13. Effects of differential thermal expansion between transmission line and tower.

aligned by pins prior to tightening hardware. Make sure the installers attach these pins to their tool pouches using a long, strong cord. This will prevent the pins from inadvertently being dropped and, subsequently, not used. Misaligned flanges will cause power to be reflected at the point, adding to system VSWR and reducing transmission efficiency.

Pressurization

Positive dry air or nitrogen pressure should be maintained inside transmission lines to prevent breathing. Ambient air can enter the line through the joints due to changes in atmospheric pressure. As the temperature drops, the moisture will condense on the internal surfaces of the waveguide. The accumulation of enough moisture in the waveguide can cause large changes in performance by increasing attenuation, system return loss, or both.

Rectangular waveguide, due to its shape, is usually not pressurized to levels greater than 0.25 psig to prevent distorting or bulging between the flanges. Circular and truncated waveguides, on the other hand, can be pressurized at levels from 0.5 to 2 psig thereby reducing the likelihood of moisture penetration from the breathing process.

Due to the large volume of enclosed air and large surface area of the guide, sudden changes in ambient temperature and/or solar heating can cause fast changes in internal pressure. A pressure relief value is needed to bleed off rapid increases in pressure which can occur as the rising sun warms up the waveguide. When cooling occurs (due to clouds, sunset, etc), the pressurization system must have the capacity to provide the required dry air volume quickly.

Due to the cyclic nature of waveguide pressurization systems (pressure relief followed by replacement of the purged volume), automatic dry air systems are more functional than nitrogen tank pressure systems since the latter would need regular replacement.

Testing

Electrical testing of waveguide systems should be performed after any of the following have occurred:

- Initial installation prior to applying power
- VSWR/reverse power trip
- Signal fluctuations of in-line waveform monitors or signal analyzers.

The purpose of such tests is to verify that the overall performance of the system is as good as expected. Large amounts of reflected power reduce system efficiency, create the potential for ghosting in television applications and may even cause transmitter shutdown. It is advantageous to test new systems while installation personnel are still on site. Common tests performed are VSWR/return loss sweep and RF pulse test. Network analyzers with time domain capabilities can be used for overall system performance checks and are excellent devices for locating discontinuities not detectable by RF pulse testing.

Sweep Test

The *sweep test* is a frequency domain test that measures the return loss of the transmission line over the operating channel bandwidth. This will determine how efficiently energy is being transferred from the transmitter system into the transmission line. The measurement is typically performed at the gas barrier inside the transmitter building by applying a 6 MHz swept signal to the transmission line through a directional coupler or return loss bridge. A sample of the forward signal is compared to the reflected signal and the return loss or VSWR can be calculated. If a network analyzer is used, the calculation is done automatically, and the return loss/VSWR versus frequency is displayed on a screen. This test must be done with the end of the line terminated in either an antenna or load. Note that the effects of the terminating device will be included in the measurement.

Since power levels are highest at the visual carrier frequency, most waveguide systems are optimized for return loss at that frequency. The goal is at least 32 dB return loss (1.05 VSWR) which means less than a tenth of one percent of the transmitted power will be reflected back toward the transmitter. For a 120 kW UHF application, this translates to less than 120 W of reflected power and a power transfer efficiency of greater than 99.9%.

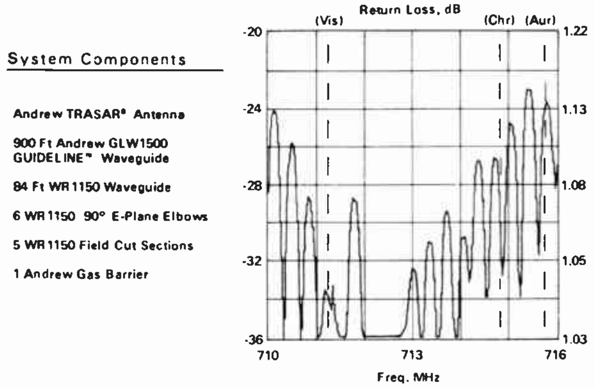


Figure 6.6-14. Sweep VSWR test results at channel 54 on 1,000 ft circular waveguide system.

A return loss sweep test of a circular waveguide system for a 240 kW application at Channel 54 is shown in Figure 6.6-14. This system included 900 ft of 15 in. circular waveguide in the vertical run, 100 ft of WR 1150 rectangular waveguide in the horizontal run, numerous bends and field cut sections, plus gas barrier and transmitting antenna. The maximum return loss level is about 23–24 dB (1.22 VSWR) at the edges of the band. At visual carrier, return loss is about 34 dB (1.04 VSWR).

The effects of discontinuities having repetitive spacing in a waveguide system, such as the flange joints, are shown clearly in Figure 6.6-15(a). The high VSWR spikes in this rectangular waveguide system occur at specific reject frequencies in the 200 MHz test bandwidth. A system utilizing different section lengths would cause these reject frequencies to change. For this reason, waveguide section lengths vary from about 11 to 12 ft depending on UHF operating channel in the same manner that rigid line lengths vary from 19 to 20 ft. Selecting the correct length insures that no VSWR spikes will occur near the 6 MHz operating band.

Rectangular waveguide systems capable of broadband operation are now being provided for special applications. Figure 6.6-15(b) shows the same waveguide system after adjusting flange tuners to reduce the effects of flange mismatch.

RF Pulse Test

An RF *pulse test* will display the performance of the line in the time domain. This test displays impedance changes versus time and allows the correlation of relative electrical performance of components to general location in the system when the propagation velocity is known. The test is performed by applying CW signal bursts of various widths to the system and detecting the reflected signals versus time with a receiver having characteristics simulating an ideal video response. A 0.25 μsec (2T) pulse has the spectrum of a normal video signal and is used to optimize the system for average picture content. The far end reflections are

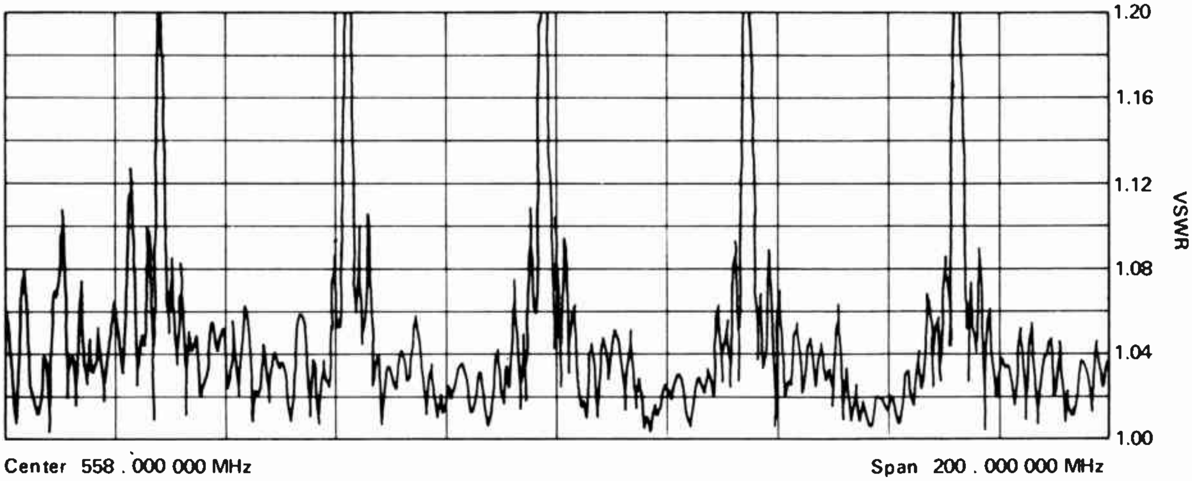


Figure 6.6-15(a). Broadband (200 MHz) test of rectangular waveguide system to show high VSWR "spikes" at frequencies corresponding to flange joint spacing.

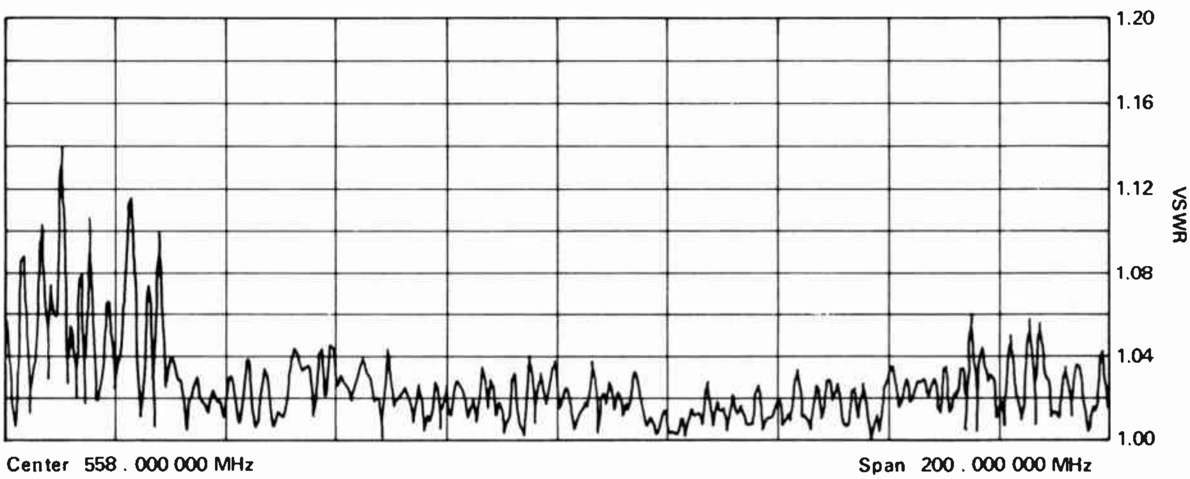


Figure 6.6-15(b). Same waveguide system optimized for broadband application using tuners at flange joints to counteract flange mismatch.

noted carefully to assure meeting the −32 dB return loss level in order to prevent noticeable ghosting. Because of the pulse width required to simulate the video information, this method does not have the positional resolution of a time-domain reflectometer (TDR). However, it provides excellent information regarding the system's transmission performance and can also detect gross problems such as misaligned flanges and pinched O-rings.

The main disadvantage of this test is that there is no standard off-the-shelf equipment package available. Therefore, correct interpretation of the test results is somewhat dependent on the skill and experience of the operator.

An RF pulse test curve for the 15 in. circular waveguide system described earlier is shown in Figure 6.6-16. As can be seen, there is nothing between the bottom gas barrier up to and including the antenna with a return loss worse than about 40 dB (1.02 VSWR) at the visual carrier frequency of 711.25 MHz.

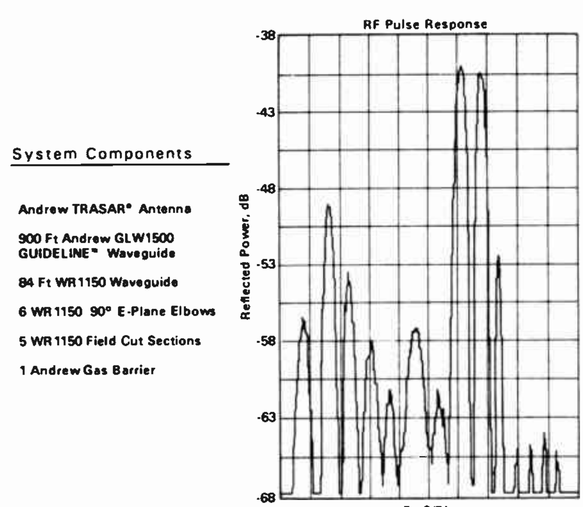


Figure 6.6-16. RF pulse test results on system shown in Figure 6.6-14.

BIBLIOGRAPHY

Andrew Corporation, "Circular Waveguide: System Planning, Installation and Tuning," Andrew Technical Bulletin 1061H, Orland Park, IL, 1980.

Ben-Dov, O., and Plummer, C., "Doubly Truncated Waveguide," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, January, 1989.

Benson, K. B., and J. C. Whitaker, "Television and Audio Handbook for Technicians and Engineers," McGraw-Hill, New York, 1989.

Cablewave Systems Technical Bulletin 21A, "The Broadcaster's Guide to Transmission Line Systems," North Haven, CT, 1976.

Carlson, Bruce, Leonard, Robert, McKee, Harry, Midkiff, John, "Transmission Lines," *NAB Engineering Handbook,* 8th ed., National Association of Broadcasters, Washington, D.C. 1992.

Crutchfield, E. B., (ed.), *NAB Engineering Handbook,* 8th ed., National Association of Broadcasters, Washington D.C., 1992.

Fink, D., and D. Christensen (eds.), *Electronics Engineer's Handbook,* 3rd ed., McGraw-Hill, New York, 1989.

Fink, D., and D. Christiansen (eds.), *Electronics Engineer's Handbook,* 2nd ed., McGraw-Hill, New York, 1982.

Jordan, Edward C., (ed.), *Reference Data for Engineers: Radio, Electronics, Computer and Communications,* 7th ed., Howard W. Sams, Indianapolis, IN, 1985.

Krohe, Gary L., "Using Circular Waveguide," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, May, 1986.

Ramo, S., Whinnery, J.R., and Van Duzer, T., Fields and Waves in Communication Electronics, Wiley, NY, 1965.

Reference Data for Radio Engineers, 6th ed., Howard W. Sams and Co., Inc., Indianapolis, IN, 1975.

Smith, P.H., Electronic Applications of the Smith Chart, Kreiger Publications, 1983.

Strickland, J., Time-Domain Reflectometry Measurements, Tektronix, Inc., Beaverton, OR, 1970.

Terman, F. E., *Radio Engineering,* McGraw-Hill, New York, 1947.

Whitaker, Jerry C., *Power Vacuum Tubes Handbook,* 2nd ed., CRC Press, Boca Raton, FL, 1998.

Whitaker, Jerry. C., *Radio Frequency Transmission Systems: Design and Operation,* McGraw-Hill, New York, 1990.

6.7
DIPLEXERS, COMBINERS AND FILTERS

WILLIAM DECORMIER
DIELECTRIC COMMUNICATIONS, RAYMOND, ME

INTRODUCTION

The components that connect, interface, transfer and filter RF energy within a given system—or between systems—are critical elements in the operation of a television station. Such hardware determines—to a large extent—the overall performance of the RF plant. In order to optimize the performance of a station's transmission system, it is necessary to first understand and optimize the components upon which the system is built.

SWITCHING

Many transmitter installations require some type of switching in the transmission line system. The complexity of this switching will depend on the individual needs of the station. It can be as simple as a three-port patch panel to connect the output of the transmitter to the station load, or as complex as a multiple transmitter antenna installation.

Patch Panels

Patch panels provide a convenient method of rerouting the interconnecting transmission lines between various inputs and outputs in the transmitter plant. Since they are manual devices and cannot be changed very rapidly, their use is limited to maintenance functions or as a secondary means of switching.

Coaxial Patch Panels

Coaxial patch panels can have any number of ports, but the more common have 3, 4 and 7 ports. They consist of the appropriate number of quick disconnect connectors mounted on a panel, and interconnecting transmission lines, usually in the form of U-lines.[1] The connectors are spaced so that the U-links may be used to interconnect any two ports (see Figure 6.7-1). For example: the connectors of a three-port patch panel would form an equilateral triangle.

Interlock switches are used to prevent transmitter power from being applied until the U-links are in the proper positions and properly seated. Power handling capabilities of the patch panels are essentially the same as the mating transmission lines.

[1] Short lengths of transmission line with connectors on both ends.

Waveguide Patch Panels

Waveguide patch panels are used in much the same way as their coaxial counterparts. However, since waveguide is larger and inflexible, and most transmitter installations use rectangular waveguide inside the transmitter building, the patch panels aren't quite as versatile. Generally, waveguide patch panels are only available as a three-port unit (see Figure 6.7-2).

Since the waveguide is rectangular, the ports must be in a straight line in either the broad wall or narrow wall plane.

Manual Coaxial Switches

Manual coaxial switches are generally available in either single-pole-double-throw (SPDT) or four-port transfer type configurations. They have two distinct advantages over the manual patch panels: ease of operation and speed of switching.

Most manual coaxial switches have either a lever or knob that is turned to change positions of the switch (see Figure 6.7-3). This can be accomplished in a few seconds, compared to minutes to change a manual patch panel. Like patch panels, manual coaxial switches have interlock switches to turn off the transmitters during switching.

Figure 6.7-1. Seven-port patch panel.

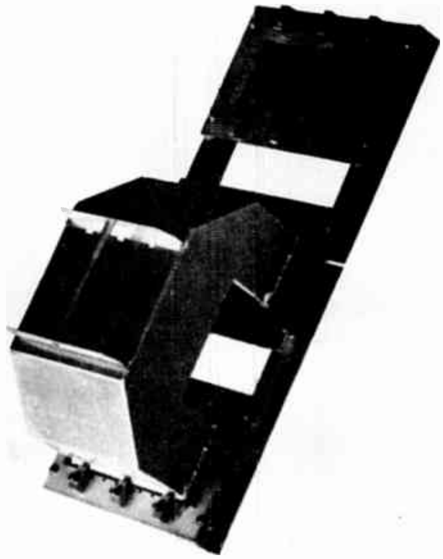


Figure 6.7-2. Three-port E-plane waveguide patch panel.

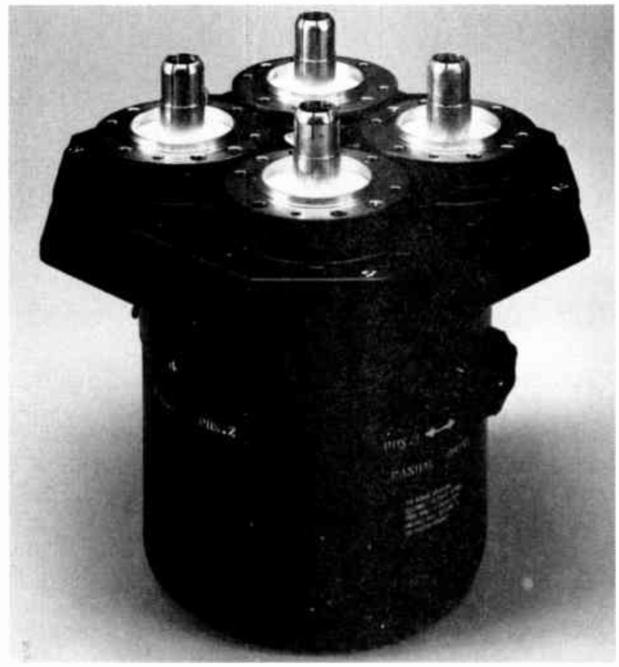


Figure 6.7-3. Coaxial rotary switch.

Power ratings of coaxial switches are approximately 80% of the equivalent coaxial transmission line. Since they are more complicated to build, they cost more than a patch panel. In order to accomplish the same functions performed by a seven-port patch panel, several switches will be required, because they are only available in three-port SPDT and four-port transfer configurations.

Motorized Coaxial Switches

Generally, motorized coaxial switches are similar to manual switches. To reduce production costs, many of the parts are the same.

Usually, the RF portion of the switch is the same for either manual or motorized switches. The knob or lever is replaced with a motor drive assembly. The motor drive system requires some type of control system to start and stop the motor in the switching sequence.

Most motorized switches are available with a choice of motor voltages, with 115 vac being the more common, and various control circuit voltages. A small control relay isolates the switch from the control circuits of the transmitter system. Generally, both sides of the control relays are available in the electrical connector. Therefore, the user must supply a power source to energize the control relay.

Rotary motorized coaxial switches are available for 1-5/8″, 3-1/8″, 4-1/16″ and 6-1/8″ cable systems. These switches will change positions in approximately two seconds. Their frequency range is good up through the TV frequencies. Some of the larger switches are limited in their use in the UHF range because of possible moding problems.

For use at higher power levels and high frequencies, another type of switch was developed. This is a motorized U-link type switch (see Figure 6.7-4). Its power ratings are essentially the same as the comparable transmission line. Sizes for of 3-1/8″, 4-1/16″, 6-1/8″, 8-3/16″, and 9-3/16″ cable systems are available. Because of the mass of the moving parts of the larger switches, switching time is increased to approximately 10 seconds.

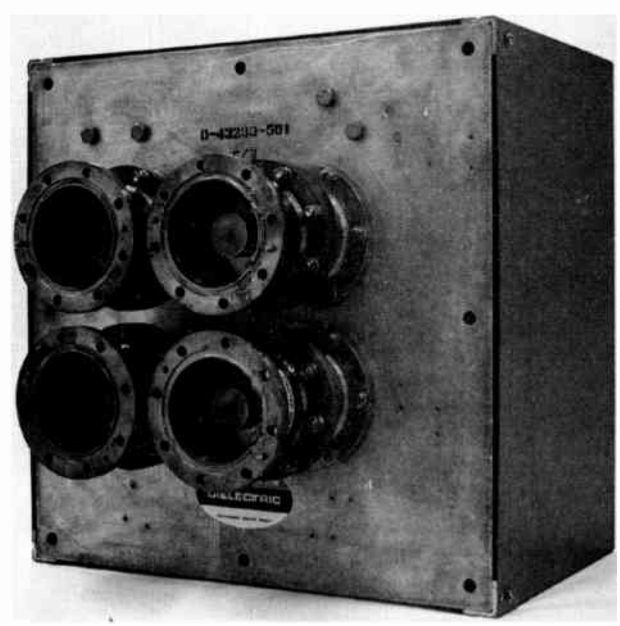


Figure 6.7-4. High power coaxial switch.

Waveguide Switches

Waveguide switches are available in three-, four-, and five-port versions. The RF sections of both manual and motorized switches are usually the same. Manual units will have a knob or handle to change the positions of the switch. Motorized units have the manual drive replaced with a motor drive assembly. Some type of motor control circuit is required.

The four-port transfer switch shown in Figure 6.7-5 is an E-plane type of switch. Both E-plane and H-plane units are available. The four-port units usually take the form of crossed waveguide, with the ports 90° apart. Within the switch is a metal back plate that has fingerstock around the four sides which contact the waveguide case. The back plate is positioned at a 45° angle to the ports of the switch. Thus, two ports of the switch will be connected together in the form of an elbow. In one position, the top port is connected to the right hand port, while the bottom port is connected to the left hand port. In the other position, the opposite is true.

The control circuits of the waveguide switch are similar to those in the coaxial switches. Various motor and control voltages are available. The waveguide switch is equipped with some type of interlock switches to turn off the transmitters during the switching process. A five-port switch is shown in Figure 6.7-6.

FILTERS

Filters are used in broadcasting to limit the undesirable emissions from transmitters. More specifically, it is necessary to limit the harmonic content of transmitters to prevent interference at higher frequencies with other services. In addition, it is necessary to limit the intrusion of other RF signals into the final stage of transmitters as a result of antenna coupling.

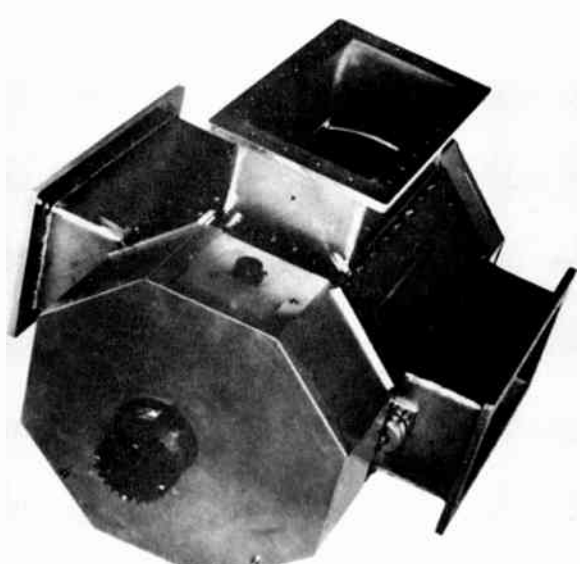


Figure 6.7-5. Four-port E-plane waveguide switch.

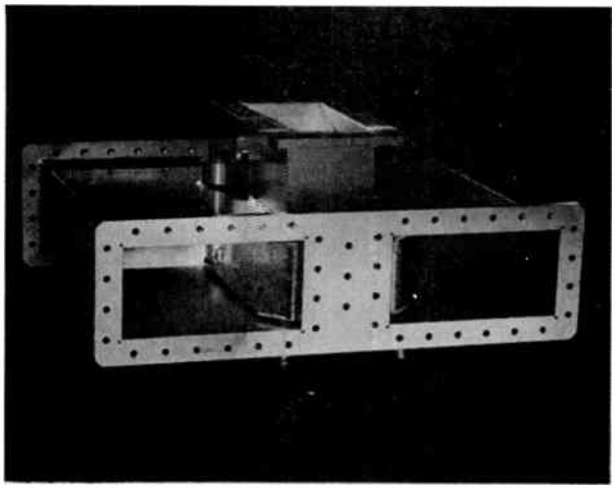


Figure 6.7-6. Five-port waveguide switch. (Courtesy Micro Communications, Inc.)

Harmonic Filters

Harmonic filters are commonly used on the output of all transmitters used for broadcast applications. They can be built for use with either coax or waveguide. The decision to use one form over the other is a matter of convenience (size), performance and cost.

The coaxial form is used in the lower portions of the broadcast spectrum, namely VHF-TV, and portions of UHF bands. For low band VHF, coaxial harmonic filters are used since the common EIA line sizes from 3-1/8 to 6-1/8 in. coaxial can handle these power levels without degradation due to higher order modes. Waveguide would have to be greater than 60 in. to operate in the fundamental mode at these frequencies.

Harmonic filters will pass the fundamental frequency with efficiencies of about 98% or -0.1 dB insertion loss. They will reject the second through the fifth harmonic with attenuation of -40 to -50 dB. By virtue of their function of attenuating harmonics they are, by necessity, designed to handle a limited segment of the band (less than one octave). Table 6.7-1 lists the typical way the band is divided.

The skirt of the attenuation curve must have a slope sufficiently large to pass the highest fundamental frequency and reject the lowest frequency in the second harmonic. A filter with nine to eleven stages will normally provide 40 to 50 dB rejection at the low end of the second harmonic.

A waveguide harmonic filter is normally used at frequencies on the high end of the UHF band. This is necessary for several reasons. Larger coaxial sizes will support the generation of higher order modes near the high end of the frequency spectrum for UHF TV. Higher order modes will be sustained at a frequency where

$$\text{Wavelength} = \pi(a + b)$$

Table 6.7-1
Filter types as a function of frequency.

Channel	Fundamental	2nd Harmonic	Typical Construction
2–3	54–66 MHz	108–132 MHz	Coaxial
4–6	66–88 MHz	132–176 MHz	Coaxial
7–13	174–216 MHz	348–432 MHz	Coaxial
14–43	470–650 MHz	940–1300 MHz	Coaxial
44–52	650–698 MHz	1300–1396 MHz	Coaxial or Waveguide
52–69	698–806 MHz	1396–1612 MHz	Waveguide

where a and b are the radii of inner and outer conductors. Larger coax sizes are also needed to handle the power levels authorized in the UHF band. 8-3/16 inch-75 Ω line will support higher order modes at frequencies just above Channel 56, and 9-3/16 inch-75 Ω cable at frequencies just above Channel 40. But the construction of a coaxial filter is a cascade of larger and smaller diameter inner conductors. The larger inner conductors essentially cause moding to occur at longer wavel engths or lower frequencies. This phenomenon lowers the effective frequency at which coaxial filters can be used.

Waveguide filter must, therefore, be used above channel 40 when transmission line power levels exceed the rating of 6-1/8″ coaxial. One type is commonly called a *waffle iron filter* because the broad walls of the waveguide resemble the top and bottom plate in a waffle iron. A waffle iron harmonic filter in WR 1150 waveguide can be used between channels 40 and 69.

Harmonic filters are usually supplied with transmitters since the transmitters cannot meet FCC Rules with regard to harmonic content without a filter. The broadcaster will seldom need to buy a harmonic filter unless he has experienced a severe transmission line failure or acquired a used transmitter.

Band-Pass and Band-Stop Filters

Band-pass filters are used sometimes in combination with band-stop filters to control another class of spurious emission problems. The transmitter (with its harmonic filter) is capable, in the absence of other RF signals, of producing transmissions which are free of any emissions. Potential problems arise, however, when the transmission facilities of two or more broadcast channels are located very close to one another.

Assume, for example, that Channel A and Channel B use antennas on the same tower. Channel A's antenna, in addition to transmitting its primary signal, also receives some of Channel B's signal. The magnitude of this received signal depends on the gain and bandwidth of Antenna A at the frequency of Channel B as well as the distance between the two antennas. Several spurious signals can be generated in Transmitter A and transmitted on the air as a result of the presence of RF from Channel B. The transmitter will usually, because of its limited bandwidth, provide several dB turn around loss for this spur. The most troublesome spur will occur at a frequency which is

$$F = 2A - B$$

A comparable problem could occur in Transmitter B where

$$F = 2B - A$$

The magnitude of the spur will be equal to the power level of the coupled signal minus the turnaround loss. So, if Channel B is present in Channel A's transmission line at a level of −40 dB down from Channel A's power level, and Transmitter A provides −10 dB turn around loss, then a spur will likely exist at (F = 2A − B) with an amplitude of −50 dB from Channel A's amplitude. In order to comply with FCC Rules for FM, therefore, a filter would have to be installed in Channel A's transmission line which would pass Channel A with minimal insertion loss (usually −0.15 dB) and provide −30 dB rejection at Channel B. This would lower the spur level to −80 dB below Channel A's transmission line level.

Another example of spurious emissions occurs when a Channel 14 or Channel 69 system is located in close proximity to land mobile communications. Two types of interference have been noted in this situation:

1. An intermodulation product of visual, aural or color signals mixing to produce a spur at a frequency used by the land mobile system.
2. The power generated by the TV station simply overloads the front end of land mobile receivers, although the carrier signals of each are at different frequencies.

The first situation must be addressed by providing filters in the TV station's system to reduce the resulting intermodulation. Figure 6.7-7 illustrates how the intermodulation was suppressed at Channel 69 in Hollywood, Florida.

The second situation must be addressed by installing a filter in the land mobile receiver system. This can generally be achieved with much smaller hardware, when required.

The decision to use band-pass versus band-stop filters should be based on the nature and extent of the problem. Each type has virtues and limitations which make them suitable for certain problems.

The typical response curves for several combinations are presented in Figures 6.7-8–6.7-11. The band-stop (as depicted in Figures 6.7-8 and 9) is characterized by rapidly rising skirts which make these filters particularly suited to rejecting frequencies that are extremely close to the desired frequency (frequencies displaced by as little as 0.8%). However, due to their sharp response, they are more prone to drifting with temperature than the band-pass configuration. They can be built so that drifting does not affect their desired performance, if manufacturers design them to minimize this tendency. Band-stop cavities are used in VHF and UHF diplexers, but are not used to protect one TV channel from another since their band-stop widths are generally too narrow to reject an entire TV channel.

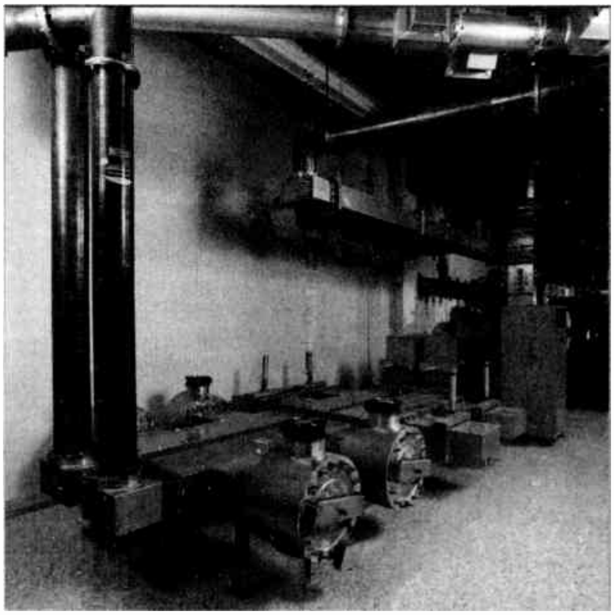


Figure 6.7-7. Notch filters at visual and aural inputs in Hollywood, Florida (Courtesy Home Shopping Network.)

The notch responses shown were plotted near 100 MHz. The center frequency can be scaled to anywhere in the VHF or UHF band, and similar percentage bandwidth would be obtainable. Figure 6.7-8 depicts the response of a single-cavity notch. Figure 6.7-9 shows a dual-cavity response.

Band-pass filter responses for a four-cavity and five-cavity filter are shown in Figures 6.7-10 and 6.7-11. Again, these were measured near 100 MHz and their percentage bandwidth would be obtainable at any other frequency in VHF or UHF spectrum. The skirt selectivity is not as sharp for band-pass cavities as for band-stop cavities. This suggests that more stages will be needed in band-pass cavities to obtain the same attenuation as a band-stop cavity. The band-pass cavity, however, has some advantages; since the pass-band is broad, it is not affected by drifting due to temperature changes. The reject curves are also located symmetrically about the pass band.

The formula for locating the spur

$$F = 2A - B$$

will always place the spur at exactly the same distance, frequency wise, from the pass-frequency, but on the

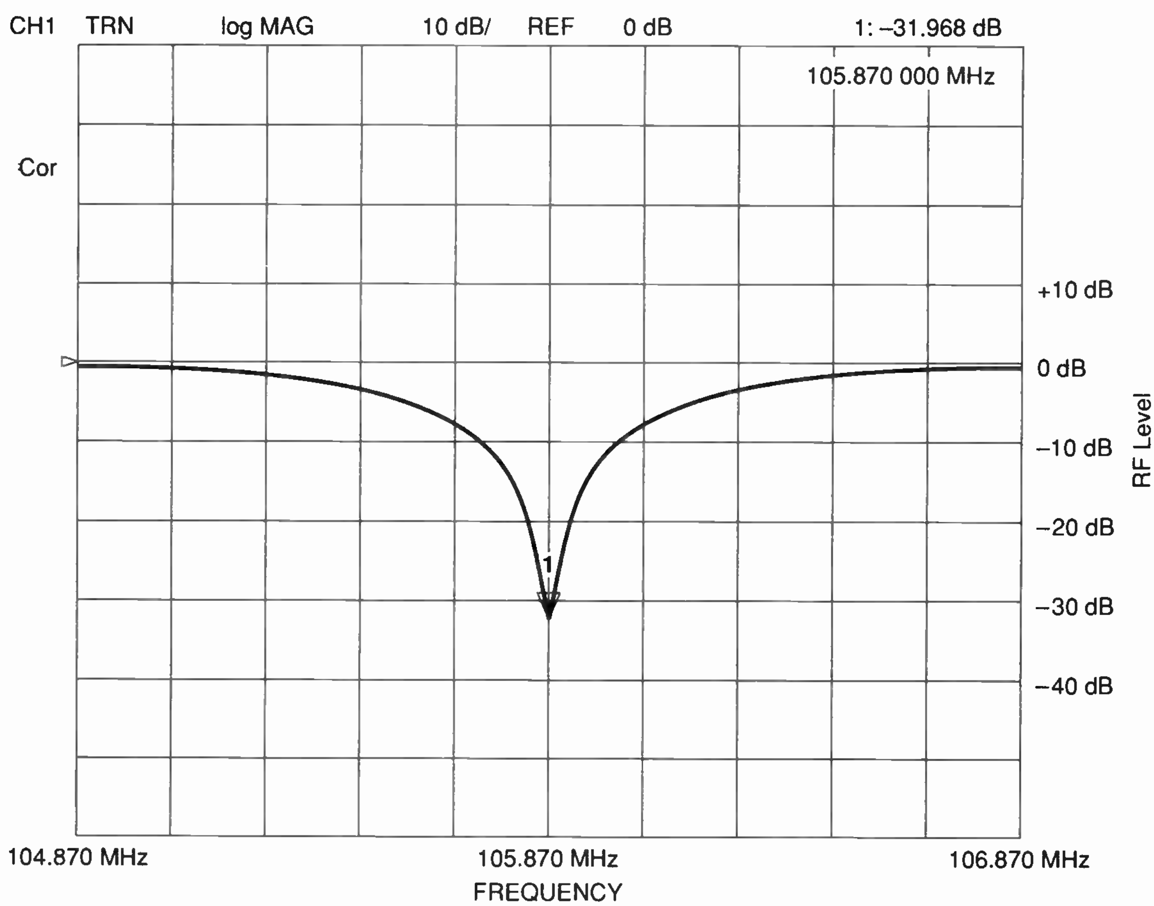


Figure 6.7-8. Single-cavity notch filter response 31.9 dB.

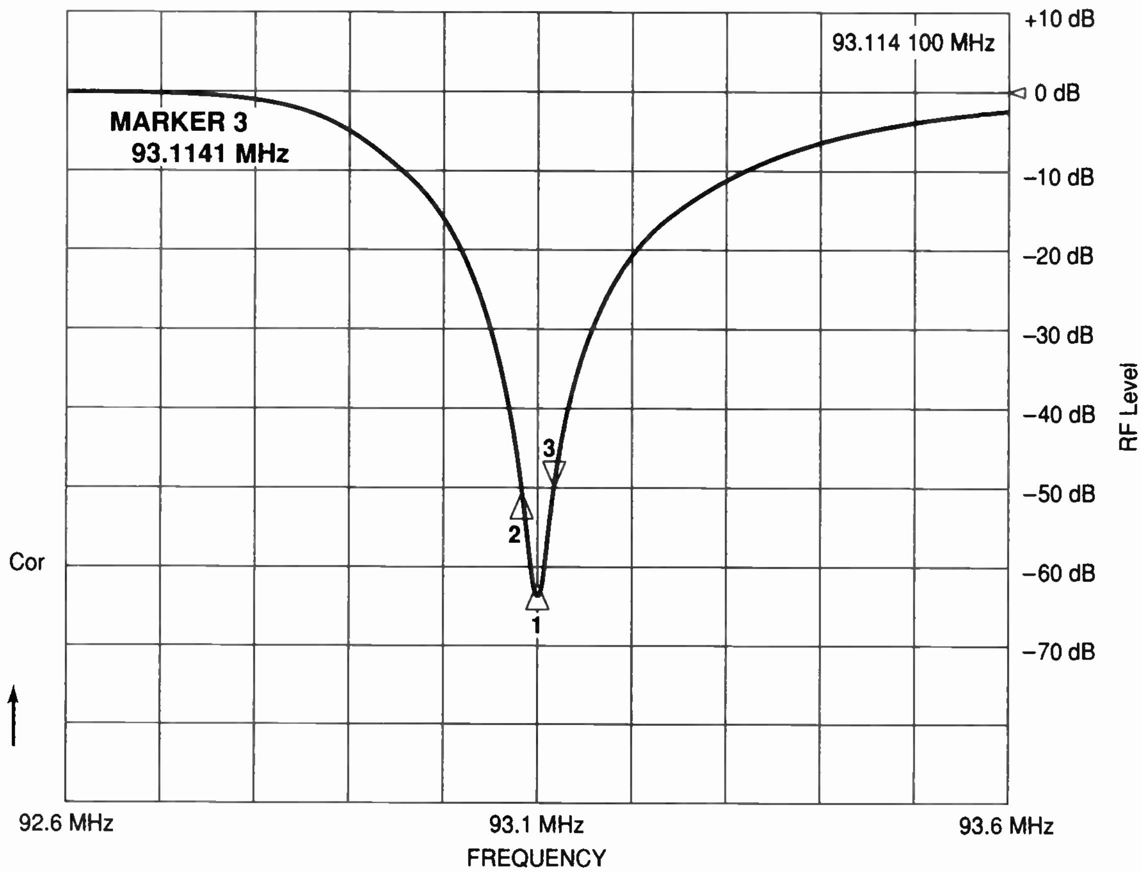


Figure 6.7-9. Dual-cavity notch filter response 64 dB.

opposite side of the pass band. So the band-pass filter attenuates the incoming RF on one side and the resulting spur on the other side of the pass band. It also has the advantage of attenuating all frequencies sufficiently removed from the pass band. This is especially useful when multiple interferences are detected.

Combinations of band-pass and band-stop are also available in a single filter. They can be visualized by superimposing a band-stop and a band-pass filter response with the notch located at the edge of the pass band. These filters are useful when interference is caused by a combination of one frequency close to the desired frequency and multiples further removed from it.

All of these responses in the VHF and FM bands can be achieved with coaxial cavities approximately 3/8 in. wavelength long and 12 to 24 in. in cross section (see Figure 6.7-12). In the UHF band they can be produced by using coaxial cavities or waveguide cavities. These filter systems also serve as building blocks for the diplexing and multiplexing systems discussed in the next section.

Once a decision is made to use band-pass, band-stop or a combination of both filters for a particular application, it is important to specify a few additional parameters to insure that the filter does not degrade to the audio or visual content of the broadcast signal.

For example, if a filter is used to prevent Channel B from entering the transmitter of Channel A, the filter will pass Channel A and reject Channel B. But the reject curve for Channel B must not infringe upon the bandwidth of Channel A. Therefore, an insertion loss variation must be specified across the operating bandwidth and perhaps beyond. At FM, where these filters are commonly used, insertion loss variation can be kept within

$$<0.1 \text{ dB}, \pm 200 \text{ kHz}$$

when either the notch is sufficiently removed in frequency or the pass band is broad enough. The edge of the reject skirt in either configuration is also characterized by a large deviation in group delay. Group delay is defined as a change in phase divided by a change in frequency.

$$\text{Group delay} = \frac{\text{Change in phase}}{\text{Change in frequency}}$$

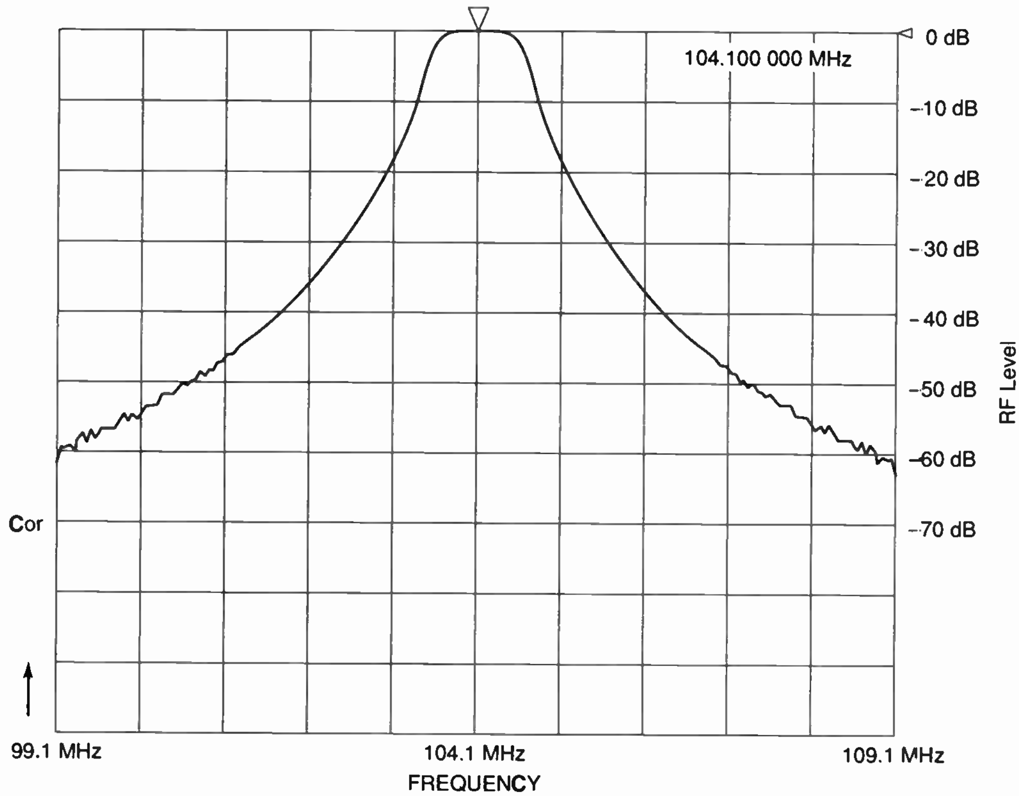


Figure 6.7-10. Four-cavity band-pass filter response.

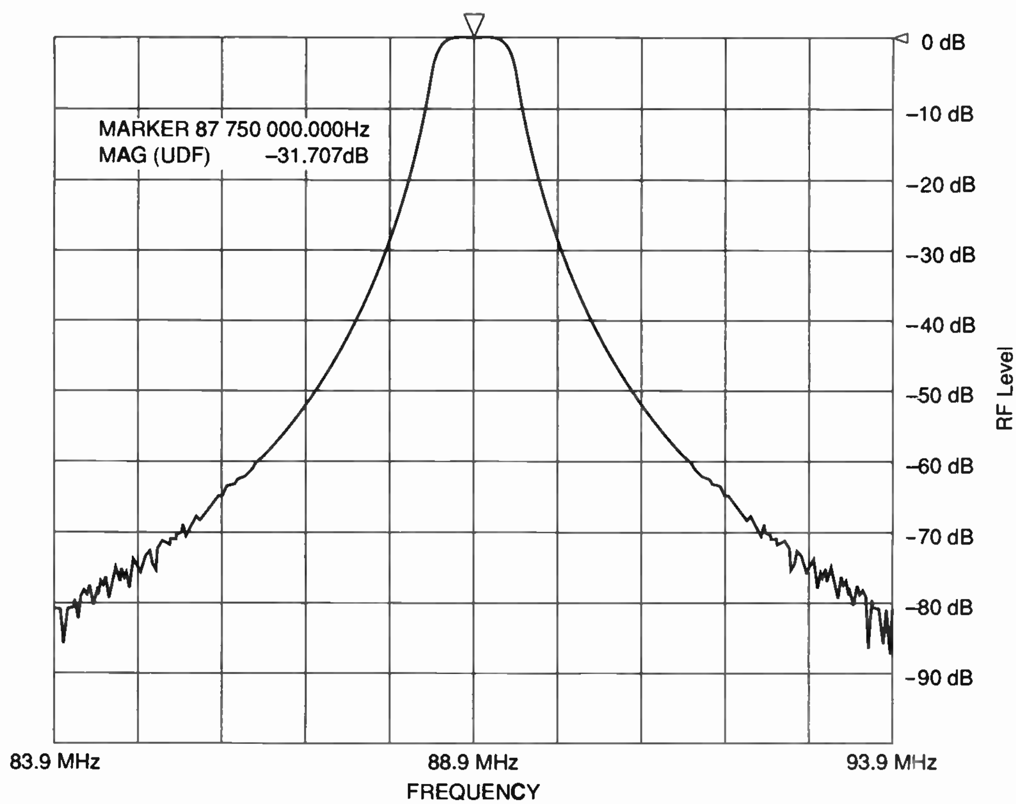


Figure 6.7-11. Five-cavity band-pass filter response.

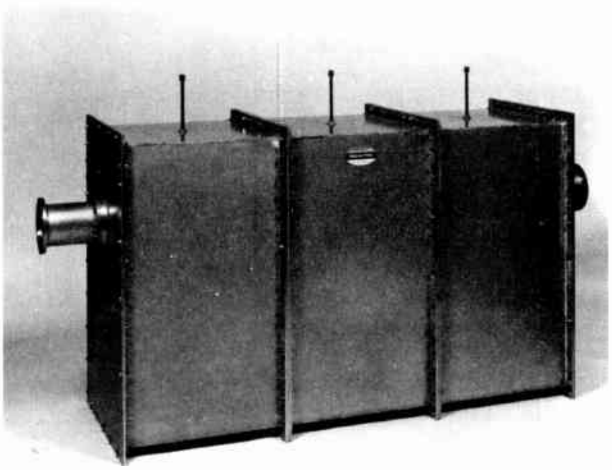


Figure 6.7-12. Three-cavity band-pass filter.

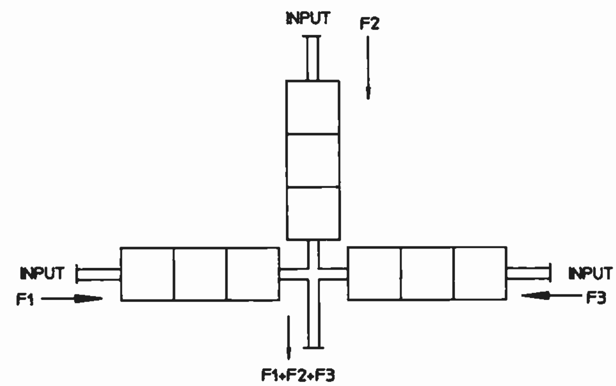


Figure 6.7-13. Schematic tee diplexer.

Group Delay Correctors

The group delay response of a band-pass filter is U-shaped. A device which uses a hybrid in conjunction with two notch cavities produces a group delay response shaped like an inverted U. Since group delay is additive, the positive group delay of the band-pass added to the negative group delay of this corrector causes cancellation at the band edges relative to center frequency. The group delay correctors thereby produce a relatively flat group delay. These were first developed as high power devices used in multiplexers. They have since been developed for low power applications and are now used in FM applications and in the aural input of TV diplexers.

Diplexers and Multiplexers

Diplexers and multiplexers are devices which allow broadcasters to combine two or more frequencies in a common transmission line while providing the isolation needed to prevent either transmitter from generating spurious emissions. Figures 6.7-13, 6.7-14, 6.7-15, and 6.7-16 provide schematic examples of configurations which can provide, in varying degree, the necessary response.

In the tee diplexer, each input leg contains either a band pass or a band-stop filter (see Figure 6.7-13). Each of these filters is characterized by a good voltage standing wave ratio (VSWR) and low insertion loss within the pass band. The slope of the reject curve away from center frequency is completely dependent on the number of cavities in each leg. Therefore, F1 and F2 must be separated sufficiently in frequency to allow the reject skirt to reach a rejection sufficient to obtain the desired isolation. All of this isolation (other than the 3 dB split of the tee) must be provided by the filter cavities. Figures 6.7-10 and 6.7-11 show rejection to be expected with deviations from center frequency. In general, F1 and F2 must be widely separated for band-pass legs and each operating band must be narrow for band-stop legs. In addition, this configuration has limitation when multiple frequencies must be combined.

Figure 6.7-14 illustrates the branch combiner for three or more channels. This device is not generally used for inputs greater than 10 W since the VSWR's and bandwidths are not as good as those in the constant impedance units shown in Figures 6.7-15 and 6.7-16. Broadcasters with tight budgets may still consider these units if they are willing to accept some performance degradation.

Figures 6.7-15 and 6.7-16, respectively, are band-stop and band-pass versions of a constant impedance diplexing configuration in common usage today at VHF and UHF frequencies. In both configurations 3 dB hybrids are used on both ends of the system and there will be an equal number of cavities in each leg between the hybrids. With the exception of the load on the isolated port and possible differences in line size, each system is electrically symmetrical about lines running through the center, both vertically and horizontally. These systems have similar response in both coaxial and waveguide configurations.

The band-stop system of Figure 6.7-15 is typically used for visual to aural diplexers in television applications.

F1, in that case, would be aural and F2 would be visual. Aural would enter at F1, be split into both legs of the hybrid, and then be reflected by the notch cavities. The reflected signal would pass through the same hybrid as before but recombine in the output port.

The visual signal (F2) would enter the rightmost hybrid, split into the two lines with cavities, pass the cavities unattenuated and recombine into the output.

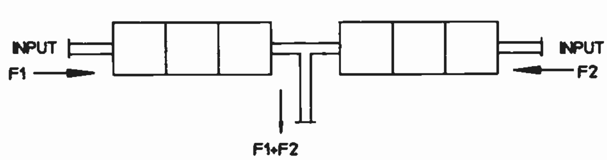


Figure 6.7-14. Schematic branch diplexer.

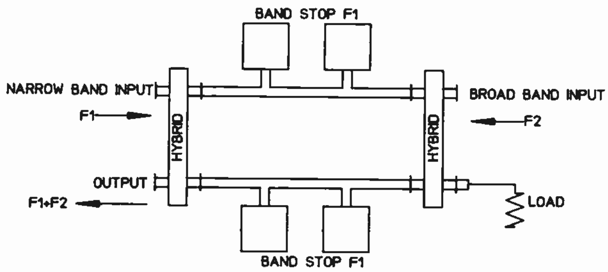


Figure 6.7-15. Schematic band-stop constant impedance diplexer.

In most TV diplexers, the second cavity in each leg would be tuned to 3.58 MHz below the visual carrier. This is the lower sideband of the visual color signal. The signal enters F2 with the visual signal and is reflected, in a manner similar to the aural signal, except that this signal is reflected into the load and absorbed.

An analysis would show the isolation of visual (F2) to aural (F1) would be due solely to the isolation inherent in the leftmost hybrid. Over a 6 MHz bandwidth, this is typically 35 dB, but can be as high as 45 dB with special care. The isolation from aural (F1) to visual (F2) is due to two components. The notches produce 27 dB typically and the rightmost hybrid about 35 dB, for a total of 62 dB.

In previous years, this type of diplexer has been used at FM frequencies, but the poor isolation in one direction required the addition of supplementary cavities on the F1 input. In addition, the group delay for this approach was four times higher than achievable with a band-pass system and the lack of symmetry and inverted shape made group delay compensation more difficult. The bandwidth, while adequate for the aural component of a TV system, has proven to be too narrow for FM applications resulting in incidental amplitude modulation.

The notch type constant-impedance diplexer has been widely used for TV applications with excellent results.

The band-pass constant-impedance diplexer, shown in Figure 6.7-16, is widely used for high-power diplexing or multiplexing applications in FM. It is occasionally used for diplexing complete TV signals, each containing visual plus aural. Several units can be cascaded together, each module providing a separate station's input. Since the cavities are band-pass units,

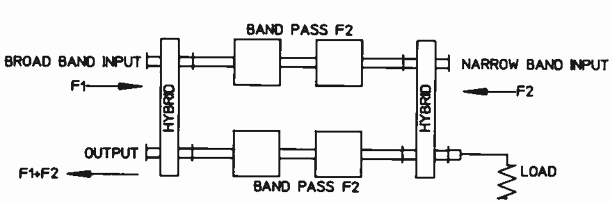


Figure 6.7-16. Schematic band-pass constant impedance diplexer.

only one channel passes through them and all others are reflected. For this case F2 becomes an individual-station input. The signal is split by the hybrid and passes with minimal attenuation through the cavities to recombine in the output. Any channel (other than F2) sufficiently spaced to be rejected by the cavities will be injected at F1, be split by the hybrid and then reflected by the cavities. The reflected signal recombines in the hybrid to exit through the output port.

An analysis of isolation in this case shows a deficiency in isolation from F2 to F1. The isolation is entirely that inherent in the hybrid. Across a 20 MHz bandwidth at FM, that isolation is typically 35 dB. The isolation from F1 to F2 is very good since it is due to the combined effects of the rejection of the cavities (which is >25 dB) and the hybrid (which is 35 dB) for a combined total of 60 dB.

If a module is used for each channel, then the adjacent module provides the additional rejection to increase the deficient isolation from F2 to F1 to 60 dB.

Figure 6.7-17 shows a four station multiplexer that is using band-pass modules to combine three FM stations with a channel 6. The taller module passes channel 6. Two supplementary notches are shown to the left of the channel 6 module to supplement the isolation from channel 6 (aural) to a closely spaced FM channel.

Figure 6.7-18 shows a VHF notch diplexer used to combine visual and aural. It functions as the schematic in Figure 6.7-15 shows.

Figure 6.7-19 shows a UHF waveguide notch diplexer. It also functions as shown in Figure 6.7-14. Figure 6.7-20 shows a UHF low power TV notch diplexer while Figure 6.7-21 is a photograph of a high band low power TV notch diplexer.

In the UHF portion of the band, where the visual transmitter is capable of amplifying the audio if diplexed low level, special motor-driven devices may be used in waveguides to detune the notch so that the entire content of the TV channel can be amplified by the visual transmitter, fed into the broadband port and recombined into the output without being attenuated by the notches.

TRANSMITTER SYSTEMS

In the following section, several methods of connecting one or more transmitters to the antenna system will be discussed. It would be impossible to cover all of the possible combinations. Therefore, only some of the basic configurations will be discussed. These can be modified or expanded to suit the individual station's requirements. Most broadcast equipment manufacturers will be happy to assist in designing custom systems.

A Caution About Interlocks

When designing an RF output system with motorized coaxial switches, it is essential to make sure that the transmitter cannot produce RF power when the switch contacts open. Most coaxial switches are con-

Figure 6.7-17. Four-channel multiplexer: Channel 6, plus three FM's. Corpus Christie, Texas.

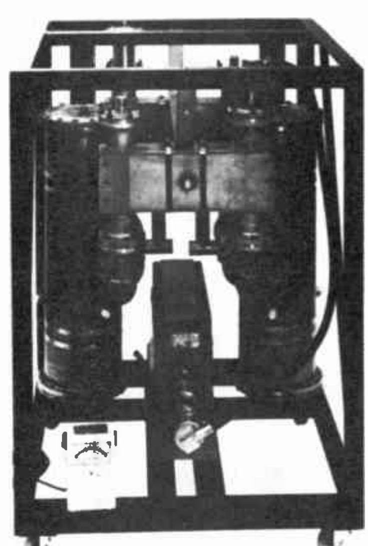


Figure 6.7-18. High band VHF notch diplexer.

structed so that the interlock kills the transmitter before the RF contacts open. However, this timing will vary between different types of switches. Also, the time it takes a transmitter to stop producing RF in response to the interlock signal will vary. Therefore,it is a good idea to check this timing. It may be necessary to turn off the transmitter a short time before commanding the coaxial switch to change positions. If the transmitter is still producing RF power when the RF contacts open, the contacts switch will be burned.

Single and Alternate Main Transmitter Systems

Figure 6.7-22 illustrates a single TV transmitter system utilizing a seven-port patch panel. The patch panel would allow the visual or aural component or the diplexer output to be terminated in station load. If the transmitter has multiplex capability (amplifying both the visual and aural signals in the visual amplifier), the visual amplifier could be connected directly to the antenna, bypassing the diplexer.

An alternate main transmitter with a coaxial switch

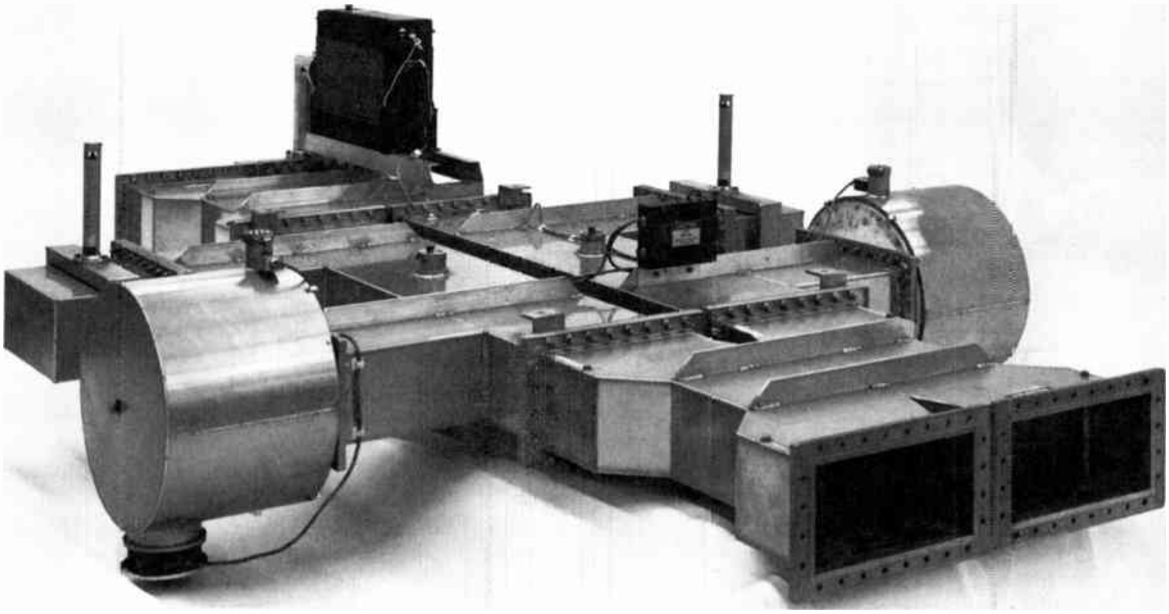


Figure 6.7-19. WR11500 waveguide notch diplexer.

is shown in Figure 6.7-23. A single ended (combined) transmitter would connect as shown. A conventional TV transmitter would require two switches.

Parallel Transmitters

Parallel transmitters are two complete transmitters whose output is combined to double the available output power. In addition to the increased output power, the parallel transmitter has additional advantages, such as redundancy. When one of the transmitters fails, the output power will drop to quarter power. If switching is provided in the output system, the output power can be increased to half power by bypassing the output combiner. This switching can take place at any convenient time after the failure.

In its simplest form, a parallel transmitter consists of an exciter modulator, input power divider, two amplifier sections, output power combiner and reject load (see Figure 6.7-24). In order to properly combine the input signals, they must be of the proper phase and amplitude. These relationships vary with the various types of combiners.

VHF Parallel Transmitters (Coaxial)

In VHF parallel transmitters, the output combiner is usually a 3 dB, 90° hybrid. In order to properly combine

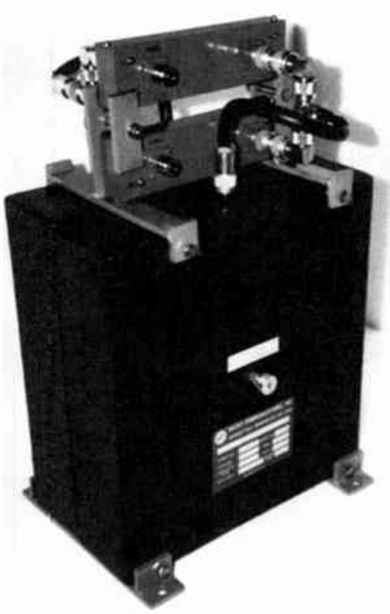


Figure 6.7-20. UHF low-power TV notch diplexer. (Courtesy Micro Communications, Inc.)

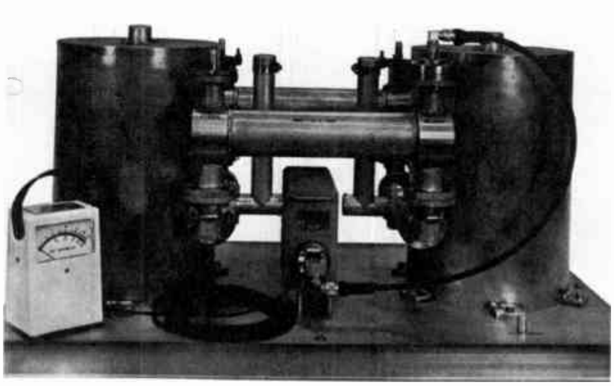


Figure 6.7-21. High band low-power TV notch diplexer.

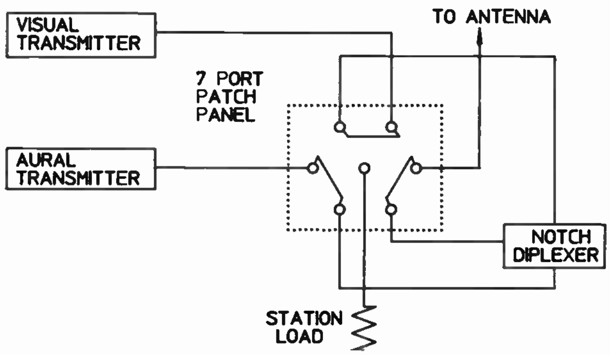


Figure 6.7-22. RF flow diagram TV transmitter with seven-port patch panel.

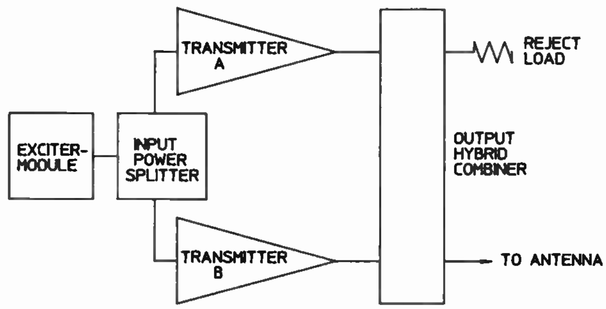


Figure 6.7-24. Basic parallel transmitter.

the two transmitter signals, the hybrid requires that the signals be of equal amplitude and phased in quadrature (90°). Assuming that the amplifier sections of the transmitters are identically tuned and that the electrical path lengths and gain are the same, then the input power divider must provide two signals that are of equal amplitude and phased in quadrature. A 3 dB hybrid will provide this type of power division, but other types could be used. An in-phase power divider with a 90° delay in one output would work just as well.

Figure 6.7-25 shows the relationship of phase error to output power of the parallel transmitter system. With no phase error (input signals to the output hybrid in quadrate) 100% of the available transmitter power will be delivered to the antenna. If there is a 90° phase error, both signals in-phase, the power will be divided equally between the antenna and reject load. All available transmitter power will be dissipated in the reject load if the phase error is 180°. A phase error of 20° will only result in an output power reduction of approximately 3%. This would indicate that the phasing is not critical for output power considerations.

The relationship between relative amplitude of the two transmitters and output power is shown in Figure 6.7-26. The graph assumes that one of the transmitters is operating at full power and that output power of the second transmitter is varied from zero to full power. If only one transmitter is operating, the output power will be only 25% of the normal combined output power. With only one input signal, the output hybrid acts as a power divider, applying half the power to the antenna and the other half to the reject load. The power being fed to the antenna is 25% of the normal combined transmitter power. If one of the transmitters is operated at half its normal output power while the other transmitter is operated at full power, the combined output power will be approximately 73% of the normal combined power. Since the two transmitters are only generating 75% of the normal combined power, only about 2% of the power is being dissipated in the reject load. Therefore, not much power is wasted in the reject load. The maximum power that the reject load should be required to dissipate is half of one transmitter's power.

Thus far, only the basic parallel transmitter system has been discussed. It was assumed earlier that the amplifier sections of the transmitters were identically tuned, having the same electrical path lengths and gain. From a practical viewpoint this could be done, but with difficulty. Therefore, most parallel transmitter systems provide a means of controlling the phase and gain of the transmitters that is independent of the tuning. This usually means an attenuator for gain and a phase shifter for the phase. Some transmitters use a gain control

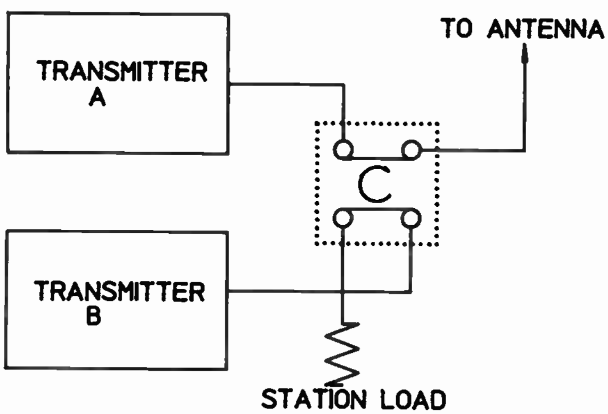


Figure 6.7-23. RF flow diagram with alternate main transmitter.

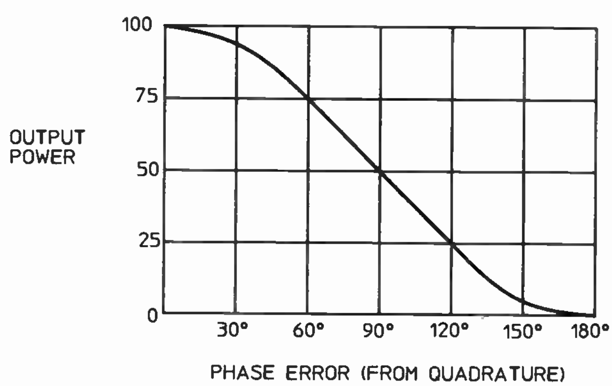


Figure 6.7-25. Relative output power versus phase error.

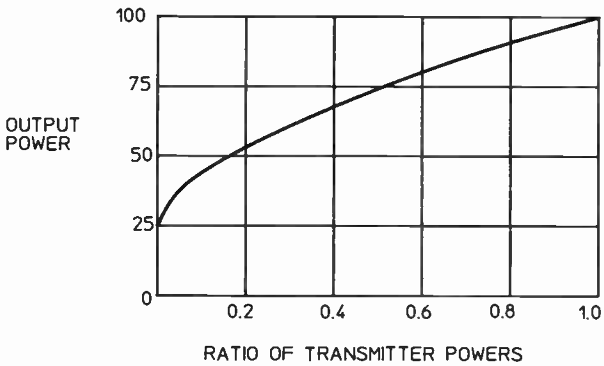


Figure 6.7-26. Relative output power versus amplitude ratio.

within the amplifier section and one of the input matching controls of the power amplifier to accomplish these adjustments. This practice is probably more common in FM than TV since the FM bandwidth is smaller.

One of the major advantages of parallel transmitters in television is the reduction of ghosts or reflections from the antenna. Ghosts occur when a portion of the power is reflected from the antenna back to the transmitter, re-reflected and finally radiated. The distance between the original image and the ghost as shown on a TV receiver can be used to determine the approximate location of the reflection in the transmitting antenna system. Since the horizontal frequency of the TV is 15,734 kHz, the full horizontal line would be equal to 63.6 μsec. The length of the visible portion of the horizontal line is 53.1 μsec. Therefore, the time between the image and the ghost can be measured. This time, when compared to the speed of light, will yield the distance the reflected signal had to travel to produce the ghost.

It must be kept in mind that the reflected signal had to travel from the point of origination down the transmission line to the transmitter and back up the transmission line to the antenna to be radiated. The velocity of propagation of the transmission line must also be taken into account.

The use of the 3 dB hybrid in the parallel transmitter will reduce the reflected signals that produce the ghosts. When the reflected energy from the antenna is applied to the output of the 3 dB hybrid, it is split into two signals at the two transmitter inputs. These signals will be phased 90° apart and will continue until they are re-reflected by the output circuitry of the transmitters. The signals will then be applied to the inputs of the 3 dB hybrid. However, their phases are such that instead of combining in the antenna output, they will combine in the reject load of the parallel transmitter. For optimum ghost reduction the electrical path lengths between the hybrid and transmitter inputs must be the same. The use of slugs or other tuning devices can upset the phase balance or electrical length of the system.

One method to measure the effectiveness of a system for ghost reduction is to measure the reverse VSWR of the system. This is accomplished by placing open or short circuits on the transmission lines that would connect to the transmitter outputs and measuring the VSWR at the output of the combining system. Thus, the path of the ghost signal is being measured. Ideally, this path should be as good as the forward VSWR of the system. However, from a practical standpoint, a VSWR of 1.1:1 or better will reduce the ghosting. It should be noted that equal electrical line lengths are needed for quadrature type combining networks. For systems using in-phase type combining networks there must be a 90° delay in the proper input. To offset this delay, a 90° delay can be inserted in the input circuitry of the opposite transmitter.

Figure 6.7-27 shows a complete output switcher for a parallel television transmitter. For single-ended transmitters such as used in FM, only half the system would be required. This figure shows the switching options of the output hybrid combiner. There are normally four modes of operation:

- A and B combined to the antenna
- A and B combined to the station load
- A to the antenna and B to the station load
- B to the antenna and A to the station load.

As drawn, the diagram is shown in the A and B to the antenna mode. By rotating S3 and S6, the system is changed to the A and B combined to the station load mode. If S1, S2, S4 and S5 are changed, Transmitter A will be connected to the antenna while Transmitter

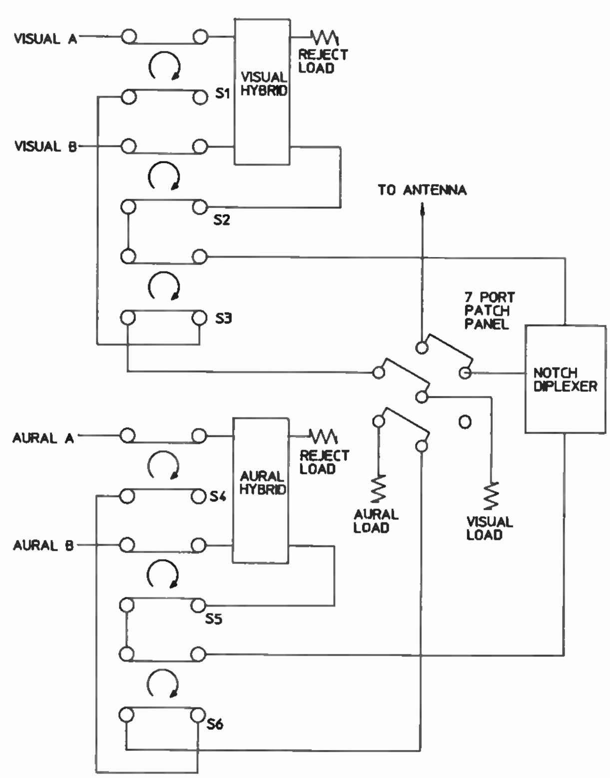


Figure 6.7-27. RF flow diagram for output switcher.

B will be connected to the station load. If S3 and S6 are rotated, the transmitters will switch Transmitter A to the station load while Transmitter B is connected to the antenna.

Since a parallel transmitter is really two complete transmitters, there should ideally be two exciter-modulators. Some manufacturers may offer the second exciter modulator as an option. By adding a switch on the input of the input power divider, either exciter modulator could be selected. This would provide redundancy, should the active exciter modulator fail. Since the power level of the exciter modulator is usually fairly low, the switching could be done under power, which would allow the switching to be automatic. By using relatively fast switches, the transfer could be done with only a small carrier interruption.

Automatic operation of the output switching system Is not usually done. Stations would rather choose when the carrier break occurs, since it will be noticeable to the audience. In lower power installations, the carrier break will be two seconds or less. With higher power switches the break could be up to ten seconds.

Again referring to Figure 6.7-27, the combined output of the visual hybrid must pass through S2 and S3. Therefore, these switches must be sized to carry the combined power. Some systems will add another coaxial switch (S7) on the output of the visual hybrid as shown in Figure 6.7-28. This higher power switch will allow the other three switches to be of a smaller size, since the combined power will only be applied to S7. It would then be necessary to add another station load for the combined signal at S7. Generally, this will

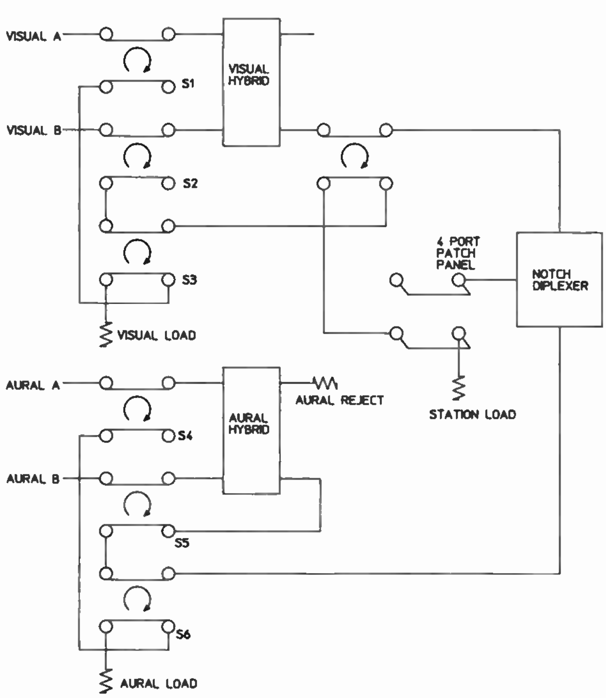
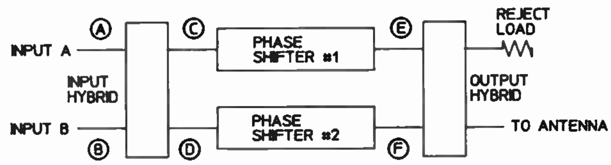


Figure 6.7-28. RF flow diagram for high power output switcher.



NOTES:

1. INPUTS A+B ARE IN PHASE
2. PHASE SHIFTER #2 WILL USUALLY CONTAIN A FIXED 90° SECTION AND A VARIABLE 90° SECTION.

MODE	PHASE SHIFTER #1	PHASE SHIFTER #2
A+B ANTENNA	0	90
A-ANTENNA B-LOAD	90	90
B-ANTENNA A-LOAD	0	180

Figure 6.7-29. Schematic of basic switchless combining system.

reduce both the physical size and the cost of the parallel transmitter system. A television system requires a diplexer of some type in order to combine the aural and visual signals. The output switching system usually contains a patch panel which allows the output of the diplexer to be routed either to the antenna or the station load. The output switching system must also contain the necessary monitoring points for combined power and reject power.

VHF Switchless System (Coaxial)

The switchless system also combines the outputs of two transmitters operating in parallel.

As with the switch method, it provides four modes of operation:

- A and B combined to the antenna
- A and B combined to the station load
- A to the antenna and B to the station load
- B to the antenna and A to the station load.

There are several methods to accomplish these mode changes without the switches. A basic system is shown in Figure 6.7-29 and consists of two 90° hybrids, a reject load, and some type of phase shifting device. The method used to shift the phases defines the various options. Therefore, the basic system will be presented here, and then the different methods of accomplishing the phase shift will be discussed.

Since the switchless system contains two hybrids, one must understand how the hybrid operates. The 90° hybrid can be used as either a power divider or a power combiner. In the switchless system both are used.

When used as a power combiner, the hybrid will combine two signals that are equal in amplitude and phased in quadrature. Again referring to Figure 6.7-29, if two equal amplitude signals are applied to points E and F, with the signal at F lagging in phase by

90°, the signals will be combined into the antenna. Conversely, if the signal at point E is lagging by 90°, the signals will be combined into the load.

If a signal is applied to the A input of the hybrid (now being used as a power divider), it will be split into two signals of equal amplitude that are phased 90° apart. The signal appearing at Point C will be in phase with the input signal (Point A), while the signal at Point D will lag the input signal by 90°.

The converse is true for a signal being fed into the B input. The signal at Point D will be in-phase, while the signal at Point C will lag 90°. If two signals that are in-phase are applied to inputs A and B, then each of the outputs (Points C and D) will have two signals, one in-phase with the inputs and one lagging by 90°.

If the switchless system is set in the A and B to the antenna mode, phase shifter #2 must have 90° more phase shift than phase shifter #1. The two signals at Point F will lag the two signals at Point E. In this situation, the signals will combine in the antenna output.

If either of the phase shifters are set for an additional 90° of phase shift, the system will be in one of the single transmitter modes of operation. If phase shifter #1 is changed, then Transmitter A will be routed to the antenna, while Transmitter B will be terminated in the station load. Should phase shifter #2 be changed, Transmitter B will be routed to the antenna, while Transmitter A is terminated in the station load.

There are three methods of sending output from the combined transmitters to the station load:

- A coaxial switch or patch panel could be used at the antenna output to route this output to a separate station load
- The combined transmitters could be routed to the reject load by moving the additional 90° phase shift in phase shifter #2 to the #1 phase shifter
- The transmitter outputs could also be combined into the reject load by changing the input phases to the switchless system. By adding 180° delay to the A input the combined transmitter output will be routed to the reject load.

The latter two methods will require that the power rating of the reject load be increased to the combined power level instead of that of a single transmitter. They also require that the antenna receive any power that is not absorbed in the reject load.

There are several methods of changing the phase of the RF signals in the switchless system.

- Probably the most familiar method of changing the phase at higher power levels is the line stretcher. It is a piece of transmission line whose length can be changed. For convenience, it often takes the form of a U-link, or trombone, so that the connectors can be mounted and the U-link moved to change the length.
- Another method uses a 90° hybrid with movable short circuits on two of the arms (see Figure 6.7-30). If a signal is applied to the input (Point A), the hybrid will divide it into two equal signals that are phased 90° apart. The signal at Point D will lag by 90°. The short circuits attached to Points C and D will reflect the two signals back to the hybrid. The phase of the two signals will be delayed by twice the electrical length of the short circuits. If the two short circuits are the same length, the relative phases of the two signals will still be 90°. Since the signal at Point D is lagging by 90°, the two signals will combine into the output (Point D). By changing the length of the short circuits, the delay or phase shift through the circuit will change. If the short circuits are moved 90°, the phase shift through the circuit will change. If the short circuits are moved 90°, the phase shift through the circuit will be 180°. The signal must travel from the hybrid to the short circuit and then return to the hybrid, twice the distance of the short circuit. The two short circuits must be moved together in order to make the hybrid combine the reflected signals properly.
- Figure 6.7-31 shows several methods of making a short circuit needed in the above circuit. It could be the traditional short (a piece of movable metal contacting the inner and outer conductors of the transmission line)
- A *deep short* could be used which moved the ringer contacts one quarter wavelength away from the short. This greatly reduces the amount of current that the fingers are required to carry. A *non-contacting short* may be used. It is a pair of cylinders that are a quarter wavelength long and shorted at one end. The sizes of the cylinders are such that they fit between the inner and outer conductors of the transmission line. The shorted cylinders are insulated from the transmission line, thus forming a capacitor. The capacitor is large enough to have very little impedance at the operating frequency, therefore, it appears as a short circuit
- Another method of creating phase shift is to use an open circuit. A short section of the center conductor is removed from a shorted piece of transmission line. The transmission line appears as an open circuit, since there is very little capacitance between the

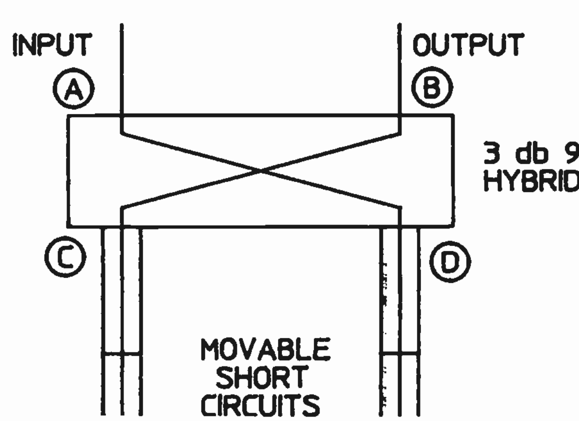


Figure 6.7-30. Variable phase shifter assembly.

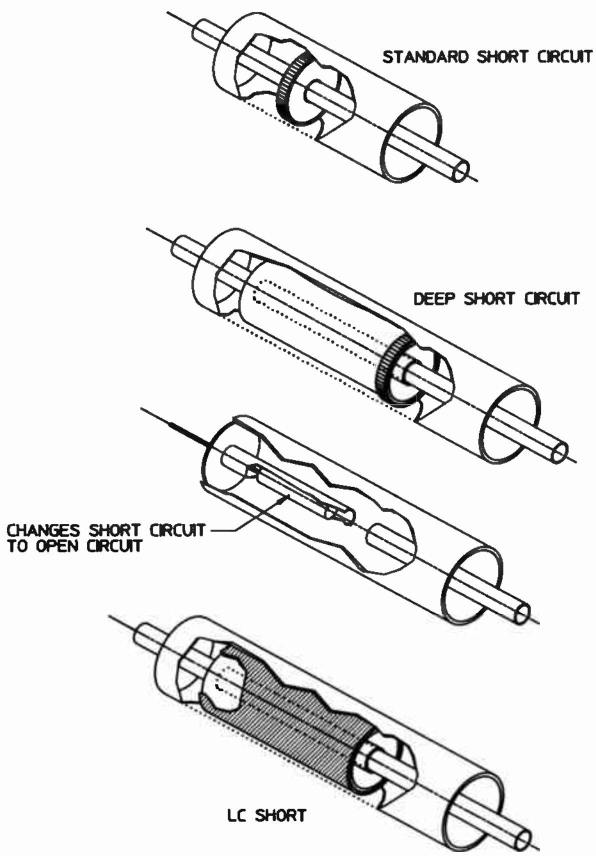


Figure 6.7-31. Variable phase shifter options.

two pieces of cut center conductor. By moving an insulated metal probe across the gap in the cut center conductor, a large amount of capacity is created. This causes the transmission line to appear as a short circuit, thus changing the phase.

All of the methods mentioned will change the phase and allow the switchless system to operate. Each method has advantages and disadvantages that should be taken into consideration upon purchasing a switchless system.

A typical flow diagram for a VHF switchless system is shown in Figure 6.7-32 and a diagram for a coaxial Opto-SX is shown in Figure 6.7-33.

The switchless system does not require that the transmitters be turned off during the switching process. Since there is no carrier break, the length of switching is not important. Switching may be done with no regard to program content.

The switchless system will offer the same ghost reduction capability as other types of parallel transmitters. Since the two signals from the transmitters are applied to the switchless system in-phase, there must be an external phase shift to take advantage of the ghost reduction feature. A 90° phase shift must be added between the output of one of the transmitters and the switchless system. This will delay the reflected signal in that path 90° as it passes from the output system to the transmitter, and another 90° as it passes from the transmitter back to the output system. The two reflected signals will now be 180° out-of-phase, and thus be combined in the reject load. Since a 90° delay was added to one of the transmitter signals, it will be necessary to add an equal phase shift to the other transmitter so that the two signals will combine in the antenna output. This delay can be added to the input circuits, allowing the ghost reduction circuit to operate properly.

Safety Considerations

There are some safety aspects of the switchless system that were not present in the switch type systems. In a switch type system, the transmitters were isolated by the mechanical switch when operating in the single transmitter mode. The switchless system does not have that isolation. Its isolation is provided only by the

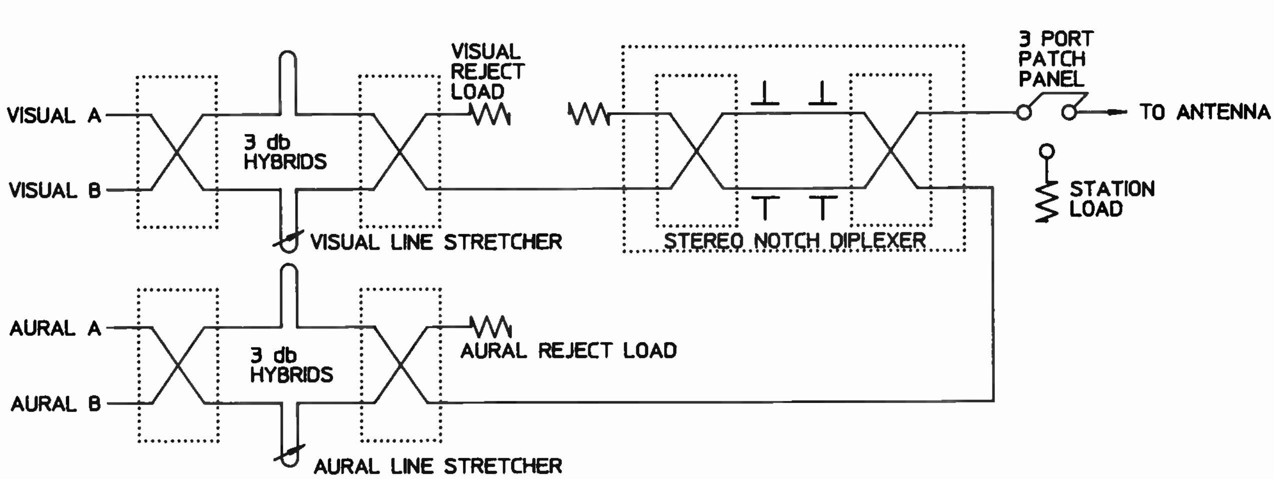


Figure 6.7-32. RF flow diagram VHF switchless system with diplexer.

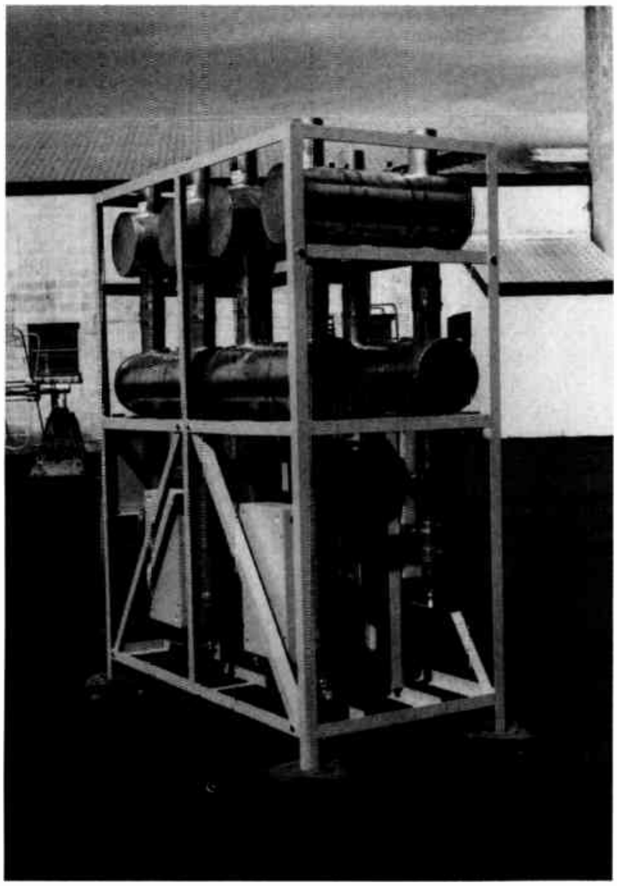


Figure 6.7-33. VHF switchless "Opto-SX" switcher. Includes aural and visua. inputs.

two hybrids. Therefore, the 60 dB or better isolation provided by the coaxial switches, is not present in the switchless system. It is possible that voltages could be present on an input, even though the transmitter driving that input is turned off. It is necessary to make sure that safety devices are used to protect technicians. It would be a good idea to delay maintenance on the system until both transmitters could be turned off.

UHF Parallel Transmitters (Waveguide)

UHF parallel transmitters are different from their VHF counterparts. Because of the power rating of the output device (typically 30 kW to 60 kW), purely parallel transmitters are not as common as in VHF. While there are quite a few 110 to 120 kW transmitters operating, these are not true parallel transmitters, as only the visual klystrons are operating in parallel. A 220 to 240 kW transmitter is the more common parallel UHF transmitter. The UHF transmitters also have aural multiplex capabilities, which reduce the need for parallel aural amplifiers.

A typical RF flow diagram for a waveguide output switching system is shown in Figure 6.7-34. It contains four waveguide transfer switches, a hybrid combiner and a reject load. S1 and S2 switch the input of the system around the hybrid for single transmitter operation. S3 determines whether the transmitter signals will be routed through the diplexer, while S4 routes the output of the diplexer to the antenna or station load. The aural amplifier is routed directly to the diplexer. This system would have the four basic modes of operation:

- Visual A and B combined to the antenna
- Visual A and B combined to the station load
- Visual A to the antenna and visual B to the station load
- Visual B to the antenna and visual A to the station load.

In addition, each of the modes could be operated either multiplexed or normal.

Recent developments in diplexers have simplified the switching systems. Aural notch detuners allow the multiplexed signals to pass through the notch diplexer rather than being switched around it. The detuners raise the aural notch cavity's frequency so that it is above the aural carrier. This allows the multiplexed visual carrier to pass through the diplexer, just as the visual signal does normally.

A station's RF output switching system is usually unique. Either the RF layout of the system will be slightly different, or the mechanical layout of the system must be adapted to the transmitter building. Since this is usually the case rather than the exception, most manufacturers are equipped to handle these situations. Figure 6.7-35 shows a system with coaxial patch panels and switches on the input and a waveguide patch panel on the output of the diplexer.

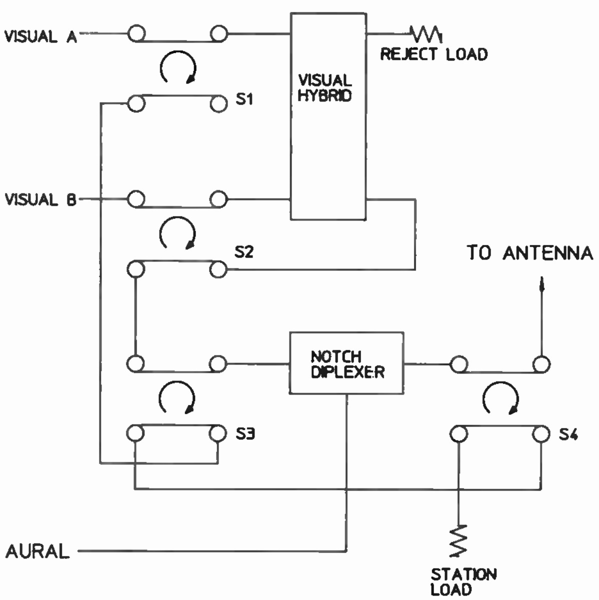


Figure 6.7-34. RF flow diagram of output switcher with parallel visual amplifiers.

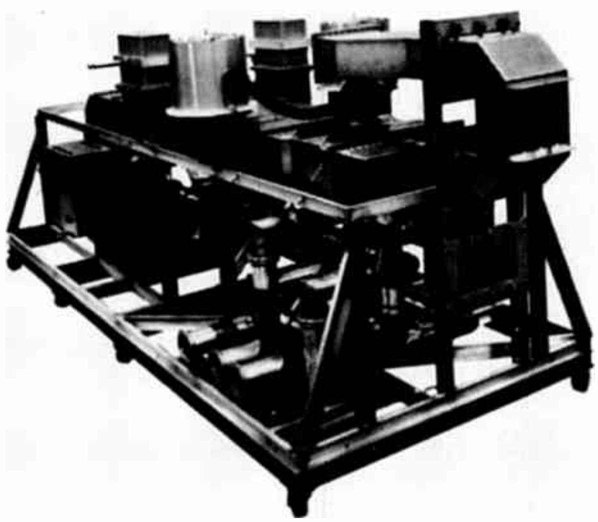


Figure 6.7-35. Output switcher 100 kW UHF transmitting system.

UHF Switchless System (Waveguide)

One of the more recent developments in RF switching is the switchless phase shifter system. These combine the outputs of higher power UHF amplifiers and can be used for either aural or visual service. This system consists of a 3 dB, 90° hybrid, a dual phase shifter, a reject load, and a magic tee (180° hybrid) and is shown in Figure 6.7-36. To better understand how the system operates, it is essential to examine the individual components.

The magic tee (180° hybrid) works much like a 90° hybrid. The difference is the amount of phase difference that it provides. When it is used as a splitter, the outputs will either be in phase or 180° out-of-phase, depending on which input is being driven. When used as a combiner, the magic tee will combine signals that are in-phase into the main output; if the signals are 180° out-of-phase it will route them to the coupled output.

Each phase shifter consists of a movable piece of dielectric material inside a section of waveguide. When the dielectric material is against the side wall of the waveguide, the phase shift will be at the minimum. As the dielectric material is moved toward the center of the waveguide, the phase shift of the signal going through the waveguide will increase. By choosing the type and amount of dielectric material, the device can be adjusted to produce 90° of phase shift. The amount of phase shift with the dielectric material against the side wall is not important in this instance. The critical element is that the phase shifts through the two units be equal.

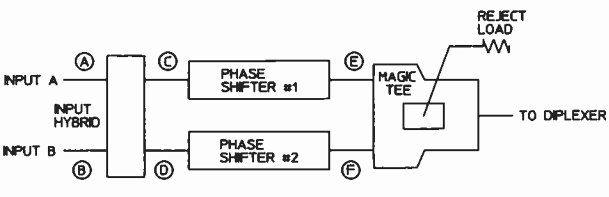


Figure 6.7-36. Block diagram UHF switchless switcher.

Assume that there is no phase shift through the unit when the dielectric material is against the side wall and 90° when it is in the center of the waveguide. If a signal is applied to the A input of the hybrid, it will be split into two signals of equal amplitude that are phased 90° apart. The signal appearing at Point C will be in-phase with the input signal (Point A), while the signal at Point D will lag the input signal by 90°. The relationship of the signals at C and D will be reversed for inputs at B. If the result of each of these scenarios is superimposed on one another to simulate the normal mode of operation with signals at both A and B, the outputs C and D each contain two signals at equal amplitude and 90° phase difference. The hybrid effectively causes these signals to be vectorially summed with an equivalent output of amplitude = 2, and phase = −45°. Although these signals have been delayed 45° with respect to the input, they are now in phase with one another at ports C and D.

With the assumption that the phase shifters have no phase shift when set to the minimum position, the signals at the output of the input hybrid (Points C and D) would be applied to the input of the magic tee (Points E and F). There will be signals present at both Points E and F that are in phase with the input signal. The signal at Point E is produced by amplifier A, while the signal at Point F is produced by amplifier B. Since these signals are equal amplitude and in-phase they will combine into the main output of the magic tee. At the same time, because the other two signals at Point E and F are equal in amplitude and in-phase, they will combine into the main output of the magic tee. At the same time, the other two signals at Point E and F are equal in amplitude and in-phase, therefore, they will combine just as the other signals did. Since these signals are not in-phase at the output of the magic tee, the resultant signal will be the vector sum.

By inserting 90° of shift in-phase shifter #1, a new mode is created. The phases at the output of the hybrid (Points C and D) will be the same as before. Since 90° is added between Points C and E, the phases at E will now be 90° for the signal from input A and 180° for the one from input B. Therefore, on the inputs of the magic tee, the signals from input A are in-phase, while the signal from the B input are 180° out-of-phase. The in-phase signals will add at the main output of the magic tee while the signals that are 180° out-of-phase will add in the coupled output.

If phase shifter #1 is returned to the minimum position and phase shifter #2 is set for 90° of phase shift, the signals from the B input will add in the main output, while the signals from the A input will add in the coupled output. By adding the phase shift in the #2 phase shifter, the signals at Point F will be 90° for the one from input B and 180° for the one from input A. Thus, the signals from input B will be in-phase at

Figure 6.7-37. UHF 120 kW switchless diplexer.

the inputs of the magic tee, while the ones from input A are 180° out-of-phase.

Thus far, three modes of operation have been accomplished within the switchless system. The fourth mode, both transmitters combined to the station load, can be done two ways. If the phase of the input signals are changed from in-phase to 180°, the combining that takes place will be applied to the coupled output (reject load) rather than to the main output. The second method is to add a switch to the output of the system to allow the output to be connected to the antenna or the station load. The latter seems to be the more common method at the present time.

The switchless phase shifter can be changed under power, since there are no contacts to break and make. The only part that is moving is the dielectric material in the phase shifter section. Since the full power of one transmitter can be applied to the reject load, the load must be capable of handling that power.

In the past, switching systems changed the RF path by mechanically moving the parts of a switch. A typical switch will have around 60 dB of isolation between the paths of the switch. In the switchless system there is no mechanical isolation. Isolation is provided by the hybrid and by the magic tee within the system. Typically, this isolation will be on the order of 35 to 45 dB, which is sufficient for good performance. However, operators must be aware that the isolation that they provide in a switch type system is not there in the switchless system. Therefore, when maintenance is being performed on the system or one transmitter, there may be a small amount of power present from the other transmitter. A typical system is shown in Figure 6.7-37.

Three Tube Switchless Combiner

The three tube switchless combiner (Dielectric, U.S. Patent #4,951,013) is another type of UHF combiner. Previously, it was not recognized that widely different amplitude of visual inputs could be combined with a switchless system. Once two tubes of a 180 kW system have been combined, the resulting 120 kW must be combined with the remaining 60 kW. The three-tube switchless enables other phase choices in addition to 0°, 90° and 180° from the phase shifters. By allowing for 19°, for example, the high efficiencies of a two-tube system may be obtained with the three tube system (see Figure 6.7-38).

This approach is also usable at VHF, should the need arise.

High Power Isolators[2]

The high power ferrite isolator offers the ability to stabilize impedance, isolate the RF generator from load discontinuities, eliminate reflections from the load and absorb harmonic and intermodulation products.[2] The isolator can also be used to switch between an antenna or load under full power, or combine two or more transmitters into a common load.

Isolators are commonly used in microwave transmitters at low power to protect the output stage from

[2] This section contributed by Jerry Whitaker, editor-in-chief.

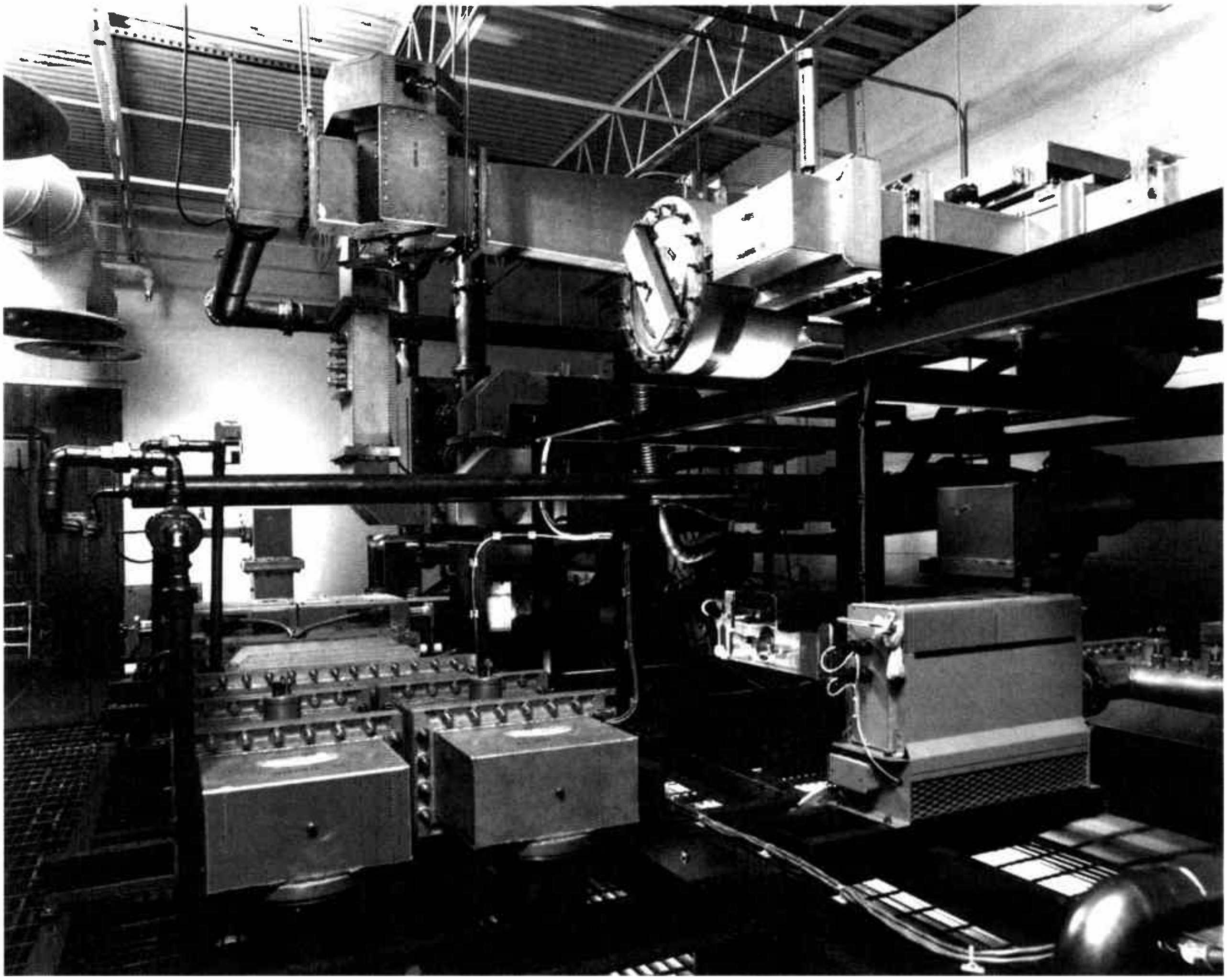


Figure 6.7-38. 180 kW three-tube switchless at channel 50 in Tampa, Florida. (Courtesy of Home Shopping Network).

reflections. Until recently, however, the insertion of the ferrite made use of isolators impractical at high power levels (10 kW and above). Ferrite isolators are now available that can handle 400 kW or more of forward power with less than 0.1 dB of forward power loss. Experimental models have been demonstrated that can handle 4 MW.

High power isolators are three-port versions of a family of devices known as *circulators.* The circulator derives its name from the fact that a signal applied to one of the input ports can travel in only one direction, as shown in Figure 6.7-39. The input port is isolated from the output port. A signal entering port 1 appears only at port 2; it does not appear at port 3 unless reflected from port 2. An important benefit of this one-way power transfer is that the input VSWR at port 1 is dependent only on the VSWR of the load placed at port 3. In most applications, this load is a resistive (dummy) load that represents a perfect load to the transmitter.

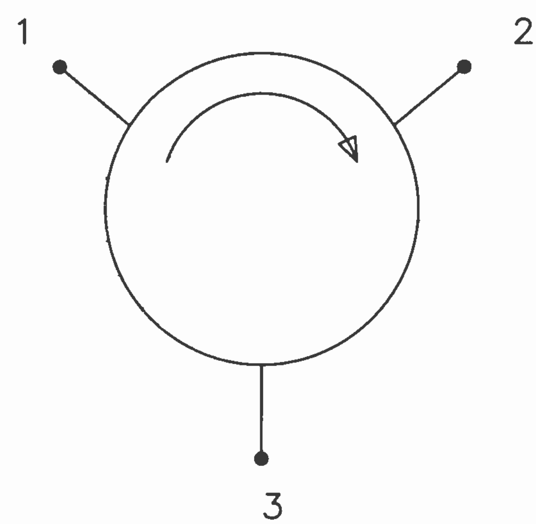


Figure 6.7-39. Basic characteristics of a circulator.

The unidirectional property of the isolator results from magnetization of a ferrite alloy inside the device. By polarizing the magnetic field of the ferrite correctly, RF energy will travel through the element in only one direction (port 1 to 2, port 2 to 3 and port 3 to 1). By reversing the polarity of the magnetic field RF flow in the opposite direction is possible. Recent developments in ferrite technology have resulted in high isolation with low insertion loss.

The ferrite is placed in the center of a Y-junction of three transmission lines, either waveguide or coax. Sections of the material are bonded to form a thin cylinder perpendicular to the electric field. Even though the insertion loss is low, the resulting power dissipated in the cylinder can be as high as 2% of the forward power. Special provisions must be made for heat removal. It is the heat removal capability that makes high power operation possible.

The insertion loss of ferrite has been kept low so that minimal heat is dissipated. Values of ferrite loss on the order of 0.05 dB have been produced. This equates to an efficiency of 98.9%. Additional losses from the transmission line and matching structure contribute slightly to loss. The overall loss is typically less than 0.1 dB, or 98% efficiency.

The ferrite element is usually water-cooled in a closed-loop system that uses an external radiator.

Applications

The high power isolator permits a transmitter to operate with high performance and reliability despite a load that is less than optimum. The problems presented by ice formations on a transmitting antenna provide a convenient example. Ice buildup will detune an antenna resulting in reflections back to the transmitter and high VSWR. If the VSWR is severe enough, transmitter power will have to be reduced to keep the system on the air. An isolator, however, permits continued operation with no degradation in signal quality. Power output is affected only to the extent of the reflected energy, which is dissipated in the resistive load.

A high power isolator can also be used to provide a stable impedance for devices that are sensitive to load variations, such as klystrons. This allows the device to be tuned for optimum performance regardless of the stability the RF components located after the isolator. Figure 6.7-40 shows the output of a klystron operating into a resistive load, and into an antenna system. The power loss is the result of an impedance difference. The periodicity of the ripple shown in the trace is a function of the distance of the reflections from the source.

Hot Switch

The circulator can be made to perform a switching function if a short circuit is placed at the output port. Under this condition, all input power will be reflected back into the third port. The use of a non-contacting high power stub on port 2 permits redirecting the output of a transmitter to the third port.

At odd quarter wave positions, the stub appears as a high impedance and has no effect on the output port.

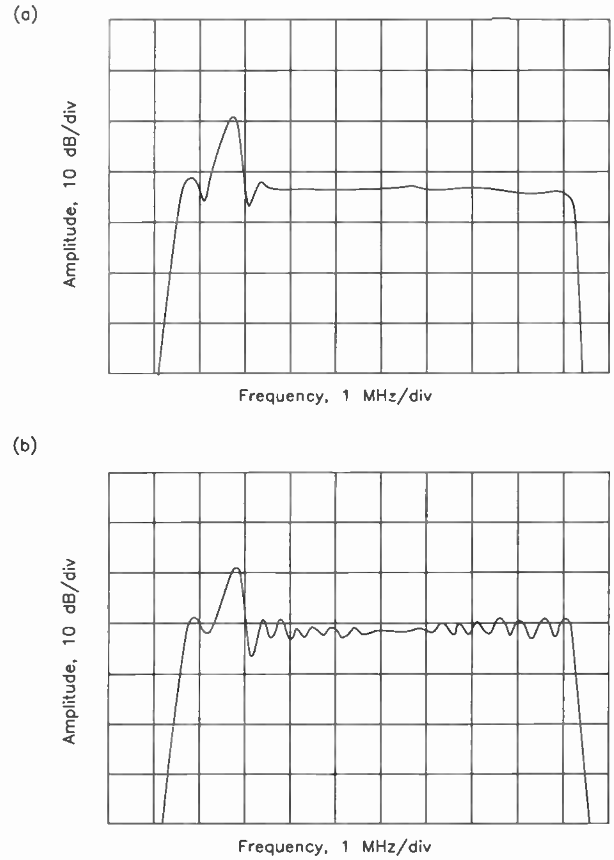


Figure 6.7-40(a). Output of a klystron operating into a resistive load, and into an antenna system (b). (Courtesy of Micro Communications, Inc.)

At even quarter wave positions, the stub appears as a short circuit. Switching between the antenna and a test load, for example, can be accomplished by moving the shorting element ¼-wavelength.

Diplexer

An isolator may be configured to combine the aural and visual outputs of a television transmitter into a single output for the antenna. The approach is shown in Figure 6.7-41. A single notch cavity at the aural

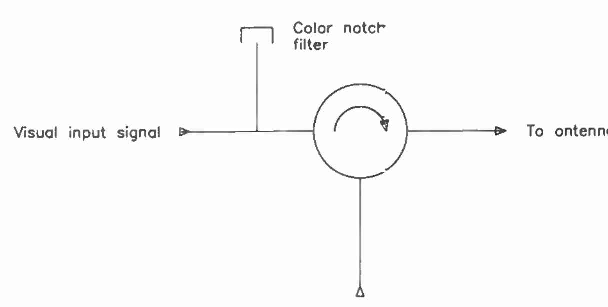


Figure 6.7-41. Use of a circulator as a diplexer in television applications. (Courtesy of Micro Communications, Inc.)

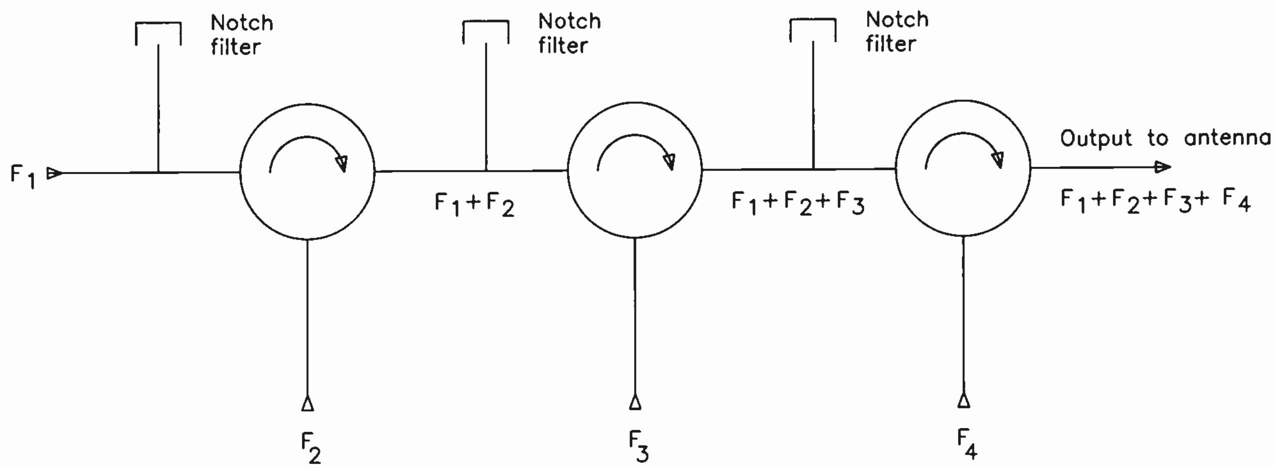


Figure 6.7-42. Using multiple circulators to form a multiplexer.

frequency is placed on the visual transmitter output (circulator input) and the aural signal is added (as shown). The aural signal will be routed to the antenna in the same manner as it is reflected (because of the hybrid action) in a conventional diplexer.

Multiplexer

A multiplexer can be formed by cascading multiple circulators, as illustrated in Figure 6.7-42. Filters must be added, as shown. The primary drawback of this approach is the increased power dissipation that occurs in circulators nearest the antenna.

CHANNEL-COMBINING CONSIDERATIONS

A number of techniques are practical to utilize an existing tower for both NTSC and DTV transmissions. Combining RF signals allows broadcasters to use an existing structure to transmit NTSC and DTV from a common line and antenna or, in the case of a VHF and UHF combination, to utilize the same line to feed two separate antennas.

During the transition period from NTSC to DTV, many broadcasters may choose to use the existing tower to transmit both NTSC and DTV channels. Some may choose to add a new line and DTV antenna; others may combine their DTV with NTSC and transmit from a common antenna and line; and still others may choose to consolidate to a new structure common to many local channels. For most stations, it is a matter of cost and feasibility.

Channel combiners (multiplexers and diplexers) suitable for DTV applications include:

- **Constant impedance.** Designs that consist of two identical filters placed between two hybrids
- **Starpoint.** Designs that consist of single bandpass filters phased into a common output tee
- **Resonant loop.** Types that utilize two coaxial lines placed between two hybrids, the coaxial lines are of a calculated length
- **Common line.** Types that use a combination of band-stop filters matched into a common output tee.

In addition, the *dual-mode channel combiner* (Micro Communications) is a device that shows promise for using a single transmission line on a tower to feed two separate antennas. Dual-mode channel combining is the process by which two channels are combined within the same transmission line, but in separate orthogonal modes of propagation.

The device combines two different TV channels from separate coaxial feedlines into a common circular waveguide. Within the circular waveguide, one channel propagates in the TE_{11} mode while the other channel propagate in the TM_{01} mode. The dual-mode channel combiner is reciprocal and, therefore, also may be used to efficiently separate two TE_{11}/TM_{01} mode-isolated channels that are propagating within the same circular waveguide into two separate coaxial lines. This provides a convenient method for combining at the transmitters and splitting at the antennas.

6.8
ANTENNAS FOR TELEVISION BROADCASTING

THOMAS J. VAUGHAN
T. VAUGHAN & ASSOCIATES, MANCHESTER, NH

INTRODUCTION

This chapter begins with a review of technical terms and FCC considerations followed by a review of antenna design considerations as they relate to these demands, as well as a review of specific applications. The chapter concludes with a review of current broadcast antennas and antenna systems.

TECHNICAL TERMS

Definitions of the following terms, consistent with FCC and industry usage, are important to the discussions in this chapter.

Effective Radiated Power

The product at the antenna input power and the antenna power gain is the effective radiated power (ERP).

Polarization

The polarization (the orientation of the electric field) can be linear horizontal, or circular. If circular is used, the term ERP applies to the horizontal and vertical components separately.

Azimuth Pattern (Horizontal Plane)

An azimuthal pattern is a plot of the free-space radiated field intensity versus azimuth angle at a specified vertical angle with respect to a horizontal plane (relative to smooth earth) passing through the center of the antenna.

A horizontal pattern is an azimuthal pattern when the specified vertical angle is zero.

For higher gain antennas where beam tilt is employed, the azimuthal pattern at the specified beam tilt is significant. In general it has been customary to determine television broadcast antenna radiation by an azimuthal pattern at the specified beam tilt and a sufficient number of vertical plane patterns taken at various frequencies in the channel.

An omnidirectional antenna is defined as one that is designed to radiate the same signal strength in all azimuthal directions. Antennas with variations up to ± 3 dB have rendered satisfactory service and are considered to be omnidirectional.

A directional antenna is one, which is designed to radiate more signal in one azimuthal direction than in other azimuthal directions.

Vertical Pattern (Elevation Plane)

The vertical pattern is a plot of free-space radiated field intensity, measured in the far field versus vertical angle in any specified azimuth plane, which contains the center of the antenna and the center of the earth.

The far field or Fraunhofer region, as usually defined, extends beyond a point where the distance between the transmitting and receiving point is

$$\frac{2a^2}{\lambda}$$

where "a" is the length of the radiating portion of the antenna and λ is the wavelength.

A requirement for the broadcast service is that a free-space radiated field should not be influenced by the proximity of the earth in such a way as to set up a nonuniform field over the antenna aperture. Proper precautions must be taken to accomplish this.

Antenna Gain

Gain is the ratio of the maximum power output at any angle from the subject antenna, to the maximum power from a thin, loss less, half wave, horizontally polarized dipole having the same power input.

Antenna gain depends on several factors, including:

- The amount of power concentrated in the maximum direction
- Losses in the antenna, which include ohmic and other losses such as energy radiated at polarizations other than the desired one.

The amount of power concentrated in the maximum direction can be determined by a comparison with a reference antenna or by integrating the total power flow through a sphere, which is done by taking a sufficient number of vertical and azimuthal patterns.

Both methods are capable of giving accurate results when the proper precautions are taken. Ohmic losses are taken into account in the comparison method or can be calculated when using the power integration method. Cross-polarized radiated energy can be measured. The measurement of gain must be carefully done with a full knowledge of all the problems that are involved. The measurement of gain used in the calculation of ERP for a circular polarized antenna must be made relative to a horizontal dipole. The directivity gain of a 1/2-wavelength dipole antenna relation to an isotopic antenna is 1.64. The gain of a circular polarized transmitting antenna relative to a like circular polarized receiving antenna will be 3 dB higher than that of a horizontally polarized dipole receiving antenna.

Gain requirements for a television broadcast antenna depend on the transmitter power, economics and field-intensity requirements as determined by the terrain and population distribution.

The maximum ERP (Effective Radiated Power) currently permitted by FCC is:

Channels 2 to 6	100 kW
Channels 7 to 13	316 kW
Channels 14 to 69	5,000 kW

Economics is a factor in antenna choice. As a general rule, for a required ERP the combined costs of transmitters and antennas are less when a higher gain antenna is used.

Impedance

Input impedance is the complex impedance looking into the antenna terminals throughout the television channel.

Most antennas are designed for the same input impedance as the standard transmission line at the antenna terminal. Impedance matching requirements for television antennas are generally more severe than for other types of service because reflected energy, which would occur when the antenna does not terminate the line properly, can cause ghosting and voltage peaks on the transmission line which could result in damage to the antenna, couplings and transmission line.

RADIATING CHARACTERISTICS

The television broadcast transmitting antenna should have an onmidirectional pattern in the azimuth plane and a narrow beam in the elevation plane. For an omni-antenna the gain is approximately one per wavelength per like polarization. Most broadcast antenna elements are spaced one wavelength apart, therefore, the gain of a linear polarized antenna is approximately equal to the number of elements.

Elevation Pattern

The elevation pattern is the product of the element pattern times the array pattern (see Figure 6.8-1).

$$F(\theta) = E_n(\tau) \sum_{n=1}^{N} A_n e^{j[(2\pi ns)\sin\tau + \delta_n]}$$

Where $E_n(\tau)$ = element elevation pattern $\approx \cos^m(\theta)$
s = element spacing
τ = elevation angle
δ_n = phase of *n*th element
A_n = amplitude of *n*th element

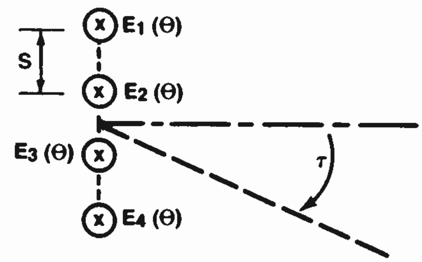


Figure 6.8-1. Elevation pattern calculation.

The resultant antenna elevation pattern for a four element array pattern spaced 1.02 wavelength with a $\cos^2(\theta)$ element pattern is shown in Figure 6.8-2.

As the spacing between the elements in the array pattern is increased, grating lobes begin to move closer to the main beam. The element pattern will no longer drive the grating lobe to zero at ±90°, resulting in high side lobes.

Figure 6.8-3 shows the patterns for a four element array with 1.30λ spacing. Although the main beam will narrow as the element spacing increases, energy will be lost in the high side lobe resulting in an over all loss of gain. The optimum spacing is one wavelength for a cos θ element pattern. Figure 6.8-4 shows power gain versus element spacing for multiple bay antennas.

Azimuth Patterns

The ripple content of the azimuth pattern is dependent on the element pattern and the distance to the phase center as illustrated in Figure 6.8-5.

$$F(\phi) = \sum_{n=1}^{N} E_n(\phi) A_n e^{j[(2\pi)\cos(\phi - \phi 2_n) + \delta_n]}$$

Where $E(\phi)$ = element pattern
ϕ = azimuth angle
δ_n = phase of *n*th element
A_n = amplitude of *n*th element
ρ = distance to phase center (in wavelengths)

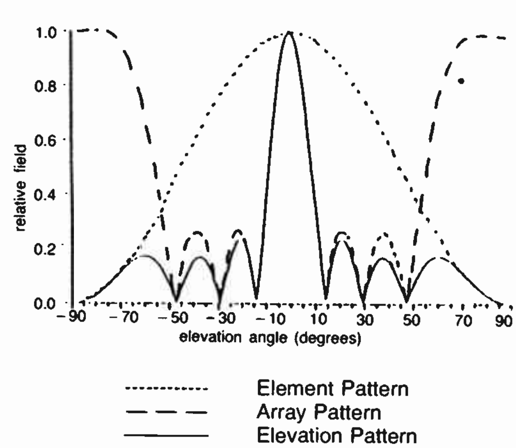


Figure 6.8-2. Elevation pattern for a four-element array with $\cos^2(\theta)$ element pattern and 1.02λ element spacing.

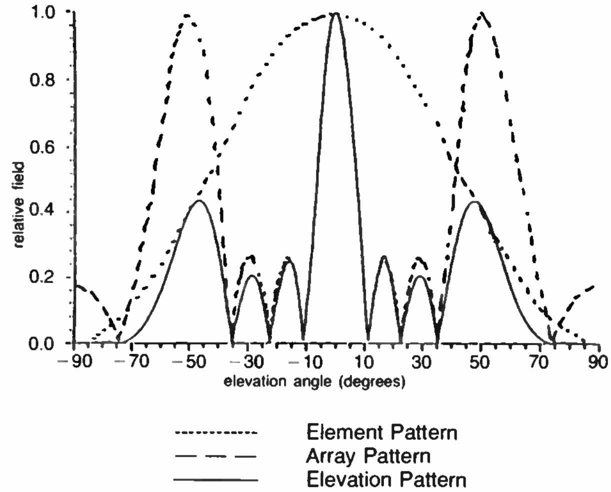


Figure 6.8-3. Elevation pattern of four-element array with $\cos^2(\theta)$ element pattern and 1.30λ element spacing.

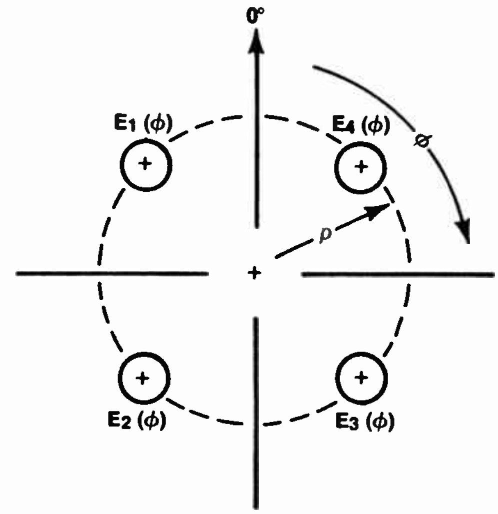


Figure 6.8-5. Azimuth pattern calculation.

The value of ρ for a panel antenna is dependent on the panel width. The value for a coaxial slotted antenna is dependent on the size of the coax line which in turn is dependent on the power handling capability of the line.

The value of ρ for waveguide slotted lines is dependent on the mode of operation which is a minimum of a half wavelength. This makes waveguide fed slot antennas practical only at UHF frequencies.

For some antennas, such as cross dipoles or turnstiles, the value of ρ is zero. Others such as panels, or slotted coax, or waveguide have a value of ρ from 0.25λ to 2.0λ as shown in Figure 6.8-6.

The ideal element pattern for a three-sided tower is $E = \cos(\theta)$; and for a four-sided tower it is $E = \cos^2(\theta)$. The ripple content of four panels with a $\cos^2(\theta)$ element pattern and phase center spacing of $\rho = 0.5, 1.5$ is shown in Figure 6.8-7.

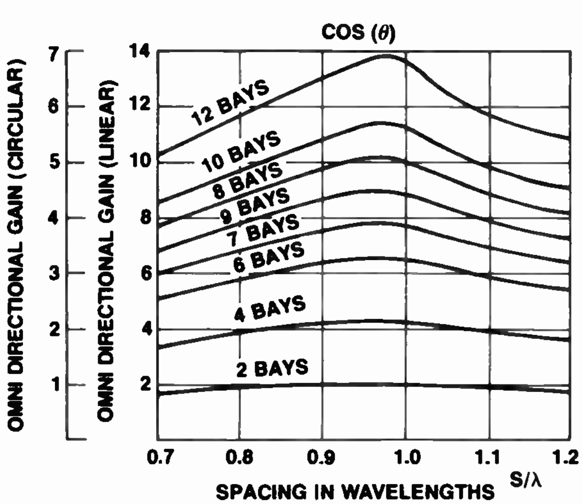


Figure 6.8-4. Power gain versus element spacing.

RADIATING STRUCTURES

There are a number of possible radiating structure designs. The most common will be discussed in this section.

Slot Antenna

The slot antenna is similar to a dipole (see Figure 6.8-8).

Currents in the slot antenna spread out over the entire sheet (in which the slot is cut) and radiation takes place from both sides of the sheet.

The resemblance between the two becomes even more pronounced when it is recognized that the field patterns of the two will be equivalent if the physical dimensions of the slot and the cross section of the dipole are the same.

Furthermore, the impedance of the slot is proportional to the admittance of the dipole of the same dimensions by the relationship

$$Z \text{ slot} = \frac{35{,}476}{Z \text{ dipole}}$$

and the bandwidth characteristics are essentially the same for both.

Actually, the previous discussion is rigorously accurate only if the sheet is of infinite extent, but it is substantially correct if the edge of the sheet is half a wavelength from the slot.

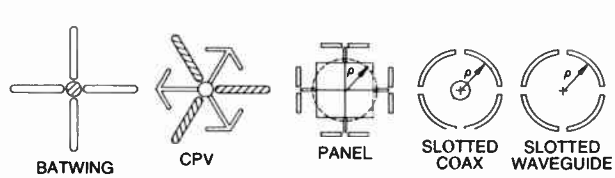


Figure 6.8-6. Azimuth pattern of four-side antenna viewed from the end.

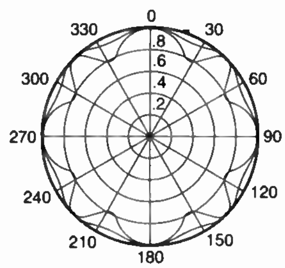


Figure 6.8-7. Arrangements of radiating elements. Azimuth versus relative field.

Bending the sheet into a cylinder results in another form of slot antenna which also takes on characteristics of a stack of coaxial rings as shown in Figure 6.8-9.

The slotted cylindrical tube can have an inner conductor and be coaxial in form or have no inner conductor and be a cylindrical waveguide.

Panel Antennas

A panel antenna has either dipoles or loops mounted $\lambda/4$ in front of a reflecting sheet. These antennas will have a directional [($\cos(\theta)$ or $\cos^2(\theta)$)] pattern in the azimuth and elevation planes depending on how it is mounted. Figure 6.8-10 shows typical panel antennas configurations.

The pattern of a panel antenna is directional since the energy radiated from the element sees its image in the ground plane resulting in a $\lambda/2$ spaced end-fire radiator. The dipoles in the *doublet panel* are spaced $\lambda/2$ and are effective in eliminating downward radiation.

Traveling Wave Antenna

Traveling wave slot antennas can be either resonant (*standing wave*) or nonresonant (*traveling wave*). Most are bottom fed. When the far end is shorted, an infinite standing wave is set up in the transmission line (either coax or waveguide). When the slots are appropriately sized and spaced, energy will be coupled out. The amplitude and phase developed across the aperture will result in a high gain antenna. The bandwidth requirements for broadcast limits this technique to narrow band or single channel applications. Several traveling wave antennas are illustrated in Figure 6.8-11.

The most common bottom fed traveling wave array is the nonresonant slot array. The number of half wave-

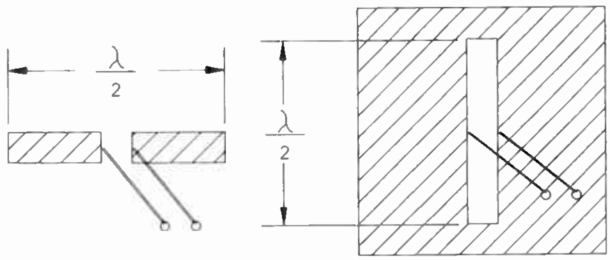


Figure 6.8-8. Dipole/slot equivalence.

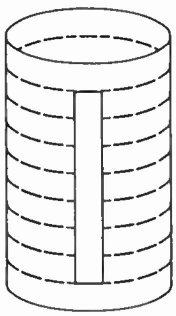


Figure 6.8-9. Slot in cylindrical sheet.

length slots arranged about the periphery is dependent on the shape of the pattern desired. The slots, circumferentially arranged around the antenna structure, are usually displaced by $\lambda/4$ and spaced vertically one wavelength apart. The slot coupling from the bottom to the top increases since the power radiated progressively reduces the power on the line. This method produces an amplitude taper across the aperture that provides null fill. With all nonresonant arrays, there is a small amount of energy leftover. This is radiated in specially designed end loaded slots or in the opposite polarization. Circular polarization can be obtained by orienting successive slots 90° and exciting them in phase quadrature, or exciting a vertically polarized radiator located near the slot.

The linear polarized spiral (helix) is essentially a strip transmission line mounted over a ground plane; the spacing controls the radiation. Each element is a $\lambda/2$ radiator and because of the orientation, the vertical components cancel. The choice of pitch angle and length along the circumference of the spiral, will result in very good horizontally polarized broadside radiation.

The circular polarized spiral or multi-arm helix consists of multiple wires mounted around a large cylinder forming a current sheet. The optimum parameters are a complex function of the number of arms, the length of one coil and the spacing to the support pole.

Excellent circularity can be obtained since in any heading the radiation is essentially that from a current ring.

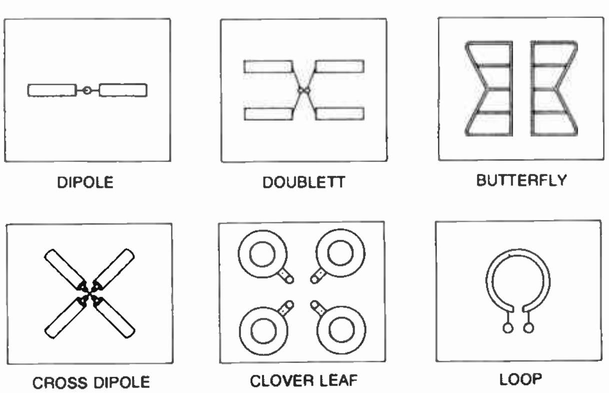


Figure 6.8-10. Panel configuration.

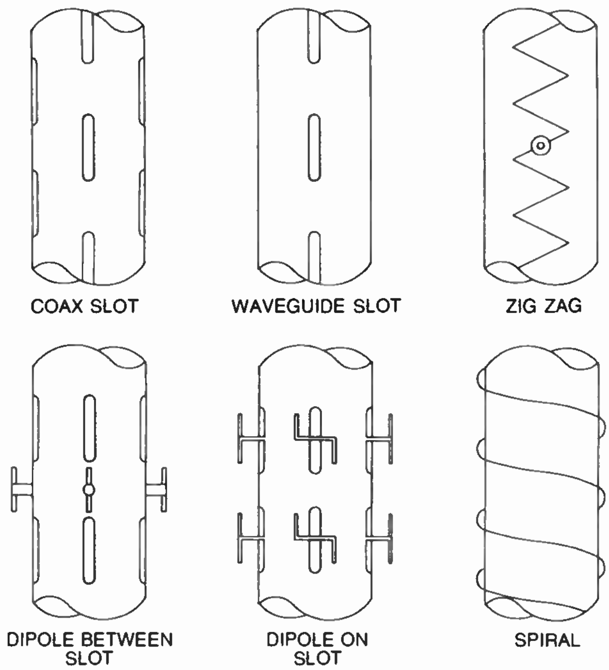
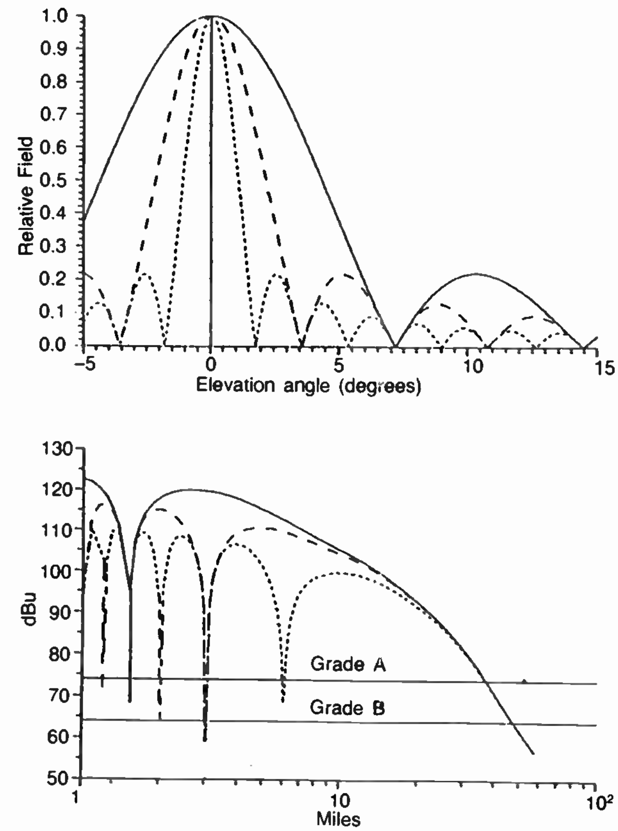


Figure 6.8-11. Traveling wave antennas.



Curve	Transmitter Power	Antenna Gain	ERP
------------	60 kW	40	2,400 kW
— — —	120 kW	20	2,400 kW
————	240 kW	10	2,400 kW

Figure 6.8-12. Signal versus antenna gain at UHF.

ERP

In determining the ERP required to serve the area under consideration, there is a trade off between using a low power transmitter and a high gain antenna, or high transmitter power and low antenna gain.

For VHF antennas, the transmitter power to antenna gain ratios are fairly well established. For channels 2 to 6, antennas usually use gain values from about 4 to 6 depending upon the length of the transmission line run. For the channels 7 to 13 band gain values vary from 12 to 18. For UHF antennas, higher antenna gains are required. Because of the much shorter wavelength at UHF frequencies, antenna with gains on the order of 20 to 30 can be built economically. At $0.10 per kW/hr, it costs $1,500/yr to generate 1 kW of RF power for a typical UHF TV station.

However, the higher gain results in narrowing the main beam. For a given transmitter input, the high gain antenna may sacrifice local coverage for more distant coverage. As the gain of the antenna increases, the beamwidth narrows. The position of the nulls and side lobes will move out from the tower.

The distance in miles from the tower is:

$$R = \frac{H}{5{,}280 \tan \theta}$$

Where:

H is the height in feet and
θ is the elevation angle

Field Strength

The elevation pattern and relative field versus distance are shown in Figure 6.8-12 for a UHF antenna on a 1,000 ft tower with constant ERP but different amounts of antenna gain.

Note that if the ERP and height above average terrain (HAAT) are held constant, there will be no change far in the far field signal strength. Figure 6.8-12 shows there is no change in relative field for the three normalized antenna patterns from 0° to 0.5°. This corresponds to a radio horizon of 20 mi.

As the gain of the antenna is reduced, the main beam will widen from 1.6° (gain 40) to 7.2° (gain 10). The null will move from 6 miles to 1.6 miles. The signal strength near the antenna will increase by 10 to 20 dB. It should be noted that this increase makes a very strong signal (100 dBμ) much stronger (120 dBμ).

Therefore, if a higher gain antenna is contemplated, the local field intensities should be calculated, using the FCC (50, 50) propagation curves. It is desirable to maintain a 100 dBμ level over the important local

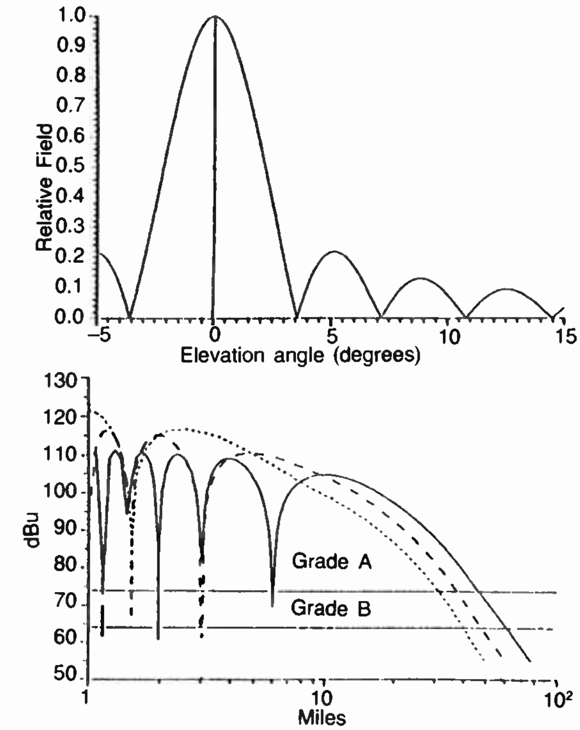
Table 6.8-1
Field intensities for specified contours.

	LoV	HiV	UHF
Channel	2–6	7–13	14–80
City Grade	74 dBμ	77 dBμ	80 dBμ
Grade A	68 dBμ	71 dBμ	74 dBμ
Grade B	47 dBμ	56 dBμ	64 dBμ

area to be covered. Most UHF antennas are designed to accomplish this with an ERP of the order of 2.5 MW at 1,000 ft. In hilly terrain, it may be desirable to increase the field by 10 dB or more; and in heavily populated cities with large building, by 6 dB or more. The field intensity specified by the FCC for the different grade levels is shown in Table 6.8-1.

If fields of this order cannot be achieved with a high gain antenna, the transmitter power should be increased.

Relative field and field intensity versus distance are shown in Figure 6.8-13 for a fixed antenna gain on a tower of 500, 1,000 and 2,000 ft.



Curve	Transmitter Power	Antenna Gain	ERP	HAAT
....	120 kW	20	2,400 kW	500 ft
— — —	120 kW	20	2,400 kW	1,000 ft
————	120 kW	20	2,400 kW	2,000 ft

Figure 6.8-13. Field intensity versus antenna height.

POLARIZATION

Polarization, illustrated in Figure 6.8-14, is defined by the plane of the electric vector (E). Elliptical polarization is the most general form. In general, an elliptically polarized wave may be expressed in terms of x and y components given by:

$$Ex = E_1 \sin(\omega t - \beta z)$$
$$Ey = E_2 \cos(\omega t - \beta z + \delta)$$

The condition for linear polarization is when either E_1 or E_2 is zero with the orientation usually horizontal.

The condition for circular polarization is when $E_1 = E_2$ and δ equals 90°, (the two components must be quadrature in both space and time).

The transmitted wave of FM and TV broadcasting signals is specified by the FCC to be linear and in the horizontal plane. AM radio on the other hand is linear and in the vertical plane. FCC regulations permit the radiation of circular polarization (CP) provided the component does not exceed the licensed ERP.

The possible advantages of using circular polarization for television are:

- Less critical receive antenna azimuth orientation; permitting good reception on all types of indoor antennas including rabbit ears, whips and rings.
- Improvement in service and penetration because the two orthogonal fields horizontal and vertical, have the same power, giving a power density twice that for a horizontal signal, alone.
- Improved coverage at the fringe area, due to twice the power density. Improvements on the order of 3 dB can be expected, moving the Grade B contour out by a calculated five to eight kilometers.

Many of the stations that changed to CP also moved their transmitting plants and changed the heights of their antennas, making a direct measured signal comparison between horizontally polarized (HP) and CP impossible. Experiences reported by stations using circular polarization indicate:

- Receiving antenna orientation is less critical, specifically for rabbit ear and whip antennas.
- Fringe area coverage has improved.
- No ill effects have taken place.
- No change in adjacent or co-channel interference has been reported.

BEAM TILT

Beam tilt is sometimes used to aim the main vertical beam tangential to the earth toward the (radio horizon), if the relative height of the antenna is such that substan-

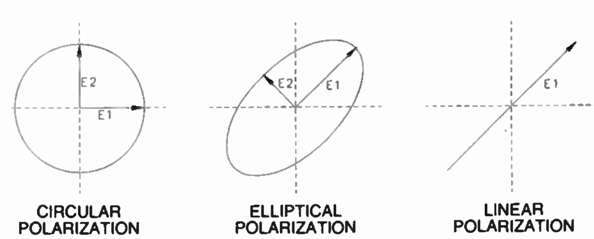


Figure 6.8-14. Polarizations.

tial energy is being radiated above the horizon. The distance to the radio and optical horizon can be determined from Figure 6.8-15.

Note that the height over the service area may not necessarily be the height over average terrain especially in mountainous areas. Also, if a body of water limits the service area, as in the case of Los Angeles, with the transmitter on nearby Mt. Wilson, it may be desirable to mechanically tilt the main beam to a point somewhat below the radio horizon.

In some cases, beam tilt may be desirable to improve local coverage. Electrical beam tilt can be used but it reduces the power gain, especially for a higher gain antenna. However, the increase in local coverage is generally a more important consideration than the slight loss of power gain. Electrical beam tilt is accomplished by delaying the energy fed to upper elements or the upper section of a split feed antenna as shown in Figure 6.8-16.

NULL FILL

The amount of null fill, and the number of nulls that need to be filled, depends upon how close the populated area is to the transmitter site. Allowance should be made for population movement towards the site in the future.

If the transmitter site is in the center of the population area or on the edge of it, consideration should be given to having null fill.

The exact amount of first and second null fill is not critical. Anything greater than 5% will usually result in signals much greater than *city grade*. The null depth (dB) in signal strength will correspond to the difference between the first null and first side lobe (dB).

The effects of null fill on signal strength is shown in Figure 6.8-17.

A combination of null fill and beam tilt is shown in Figure 6.8-18. A beam tilt of 1° will usually result in a 10 dB improvement in near-in coverage. Note that null fill and beam tilt have little effect on distant signal coverage.

POWER CAPABILITY

Power in television broadcast transmission systems is usually in terms of *peak power*, which is the instantaneous power developing during the peak of the synchronizing pulse in the visual transmitter. Since the black level signal is 0.75% of the total voltage value of the pulse, the black level power for a totally black picture is 56% of the peak-to-peak sync power. The duty cycle of the synchronizing pulses, both horizontal and vertical, adds about 4% to this power so that black level power is 60% of the peak TV power. Since the aural transmitter is usually 10% of the peak TV power, the total heating or continuous wave (CW) power of the TV signal is 66% of the peak TV power. The average power level (APL) of the video signal with typical program measures over a long period of time, 4.32 dB (37%) below the peak TV power.

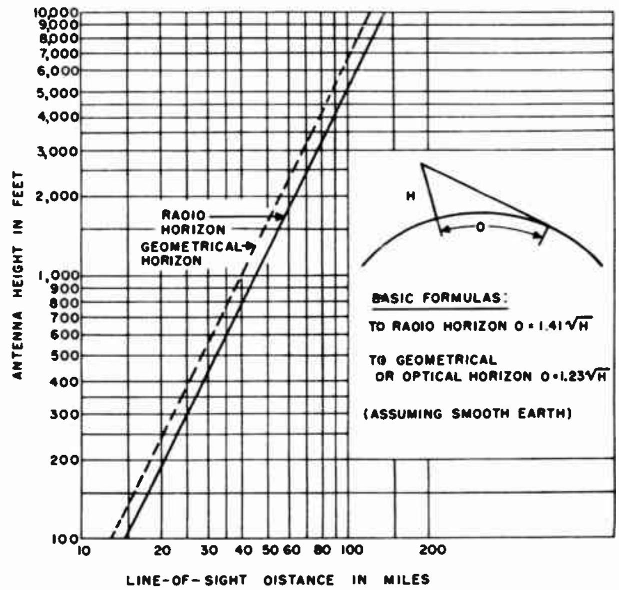


Figure 6.8-15. Distance to radio horizon versus height.

The design of all TV antennas should allow for a sufficient safety margin to handle the peak-to-peak sync level, imperfections in the transmission line, VSWR, changes in pressure and the aural power. Long transmission lines feeding the antenna will usually attenuate this figure by 10% to 20%.

Unwanted Antenna Radiation

The antenna pattern is the product of the array pattern times the element pattern. The array pattern for one wavelength spacing has a downward and upward lobe as large as the desired main lobe. If the element pattern is not a $\cos(\theta)$ function, the element pattern will not drive the array downward lobe to zero, resulting in a very large signal in the vicinity of the tower base. Figure 6.8-19 shows the resultant antenna pattern with a $\cos^{1.0}(\theta)$, where N = 0.0, 0.5 and 1.0.

Care should be exercised in making sure the element pattern has a distribution in the elevation plane of $\cos(\theta)$ or $\cos^2(\theta)$.

A well designed antenna system should not create a nonionizing radiation problem. The tower height is usually selected for maximum signal coverage and should therefore be high enough above the ground so the downward radiation is well below the safe ANSI levels.

Equally important would be the power density level on nearby towers, which could result in tower maintenance personnel being exposed to high energy levels when in the main beam of the antenna on a nearby tower.

The distance in feet along the main beam to the 1 mW/cm^2 contour for maximum ERP and the height above ground to the 1mW/cm^2 contour is listed in Table 6.8-2.

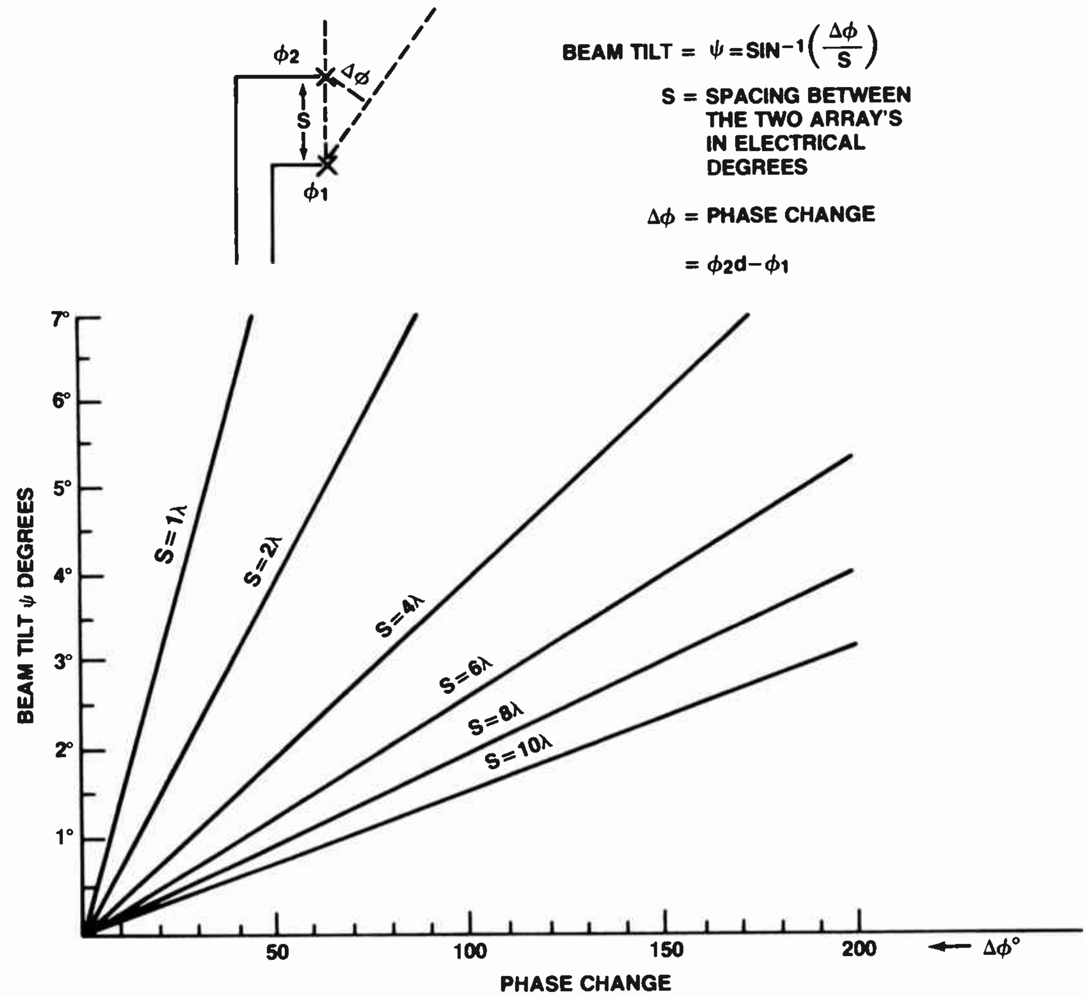


Figure 6.8-16. Beam tilt versus element phase change ($\Delta\phi$).

ANTENNA CHARACTERISTICS

The primary purpose of an input VSWR specification is to insure a good match to the transmission line. If the mismatch is a great distance away, the reflected power may be of such magnitude that it travels back to the transmitter where it is will be re-reflected back to the antenna and radiate. It will than appear as a secondary image or ghost on the television picture. The image is delayed by twice the length of the transmission line.

Subjective experiments have established that the reflection at the carrier should be no greater than 3% of the incident voltage.

Empirically derived data of objectional reflection VSWR that will cause ghosting as a function of distance in microseconds from the transmitter to the antenna is shown in Figure 6.8-20.

The dashed line represents the minimum detectable reflection readable on the window pulse.

Due to the concentration of energy, the VSWR at the picture carrier should be kept fairly low at this frequency. The values below visual carrier are not as critical since the slope in the receiver cuts off most of the energy below the carrier. The VSWR in the visual pass band should be as shown in Figure 6.8-21.

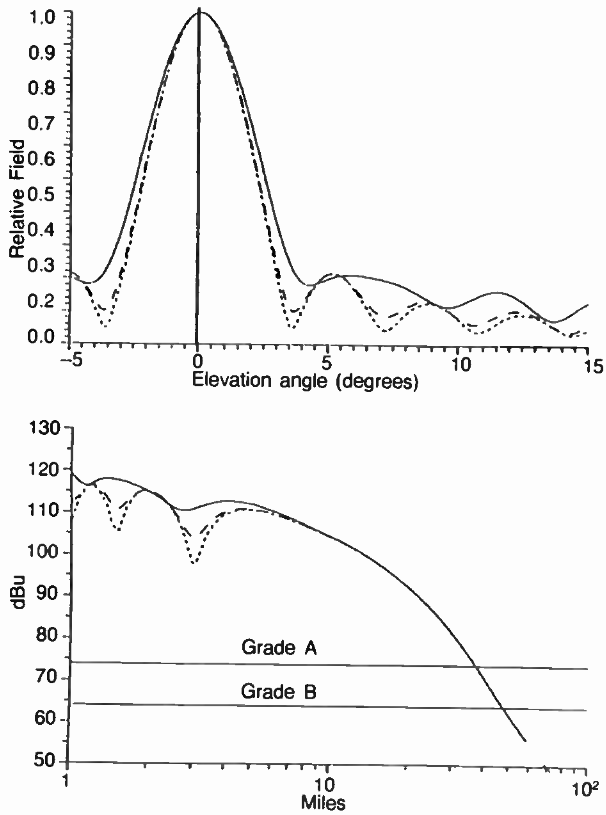
There are two locations of reflections that will cause ghosts: the lower elbow complex and the combination of the upper elbow complex and the antenna.

The upper elbow complex and the antenna reflection would be more objectionable than the lower elbow complex, because of its distance and displacement ($1/\Delta f$) rather than its magnitude. A 5% (VSWR 1.10) echo would not be as objectionable at 0.2 μsec as it would be at 2.0 μsec, as shown in Table 6.8-3.

Transmission Line and Components

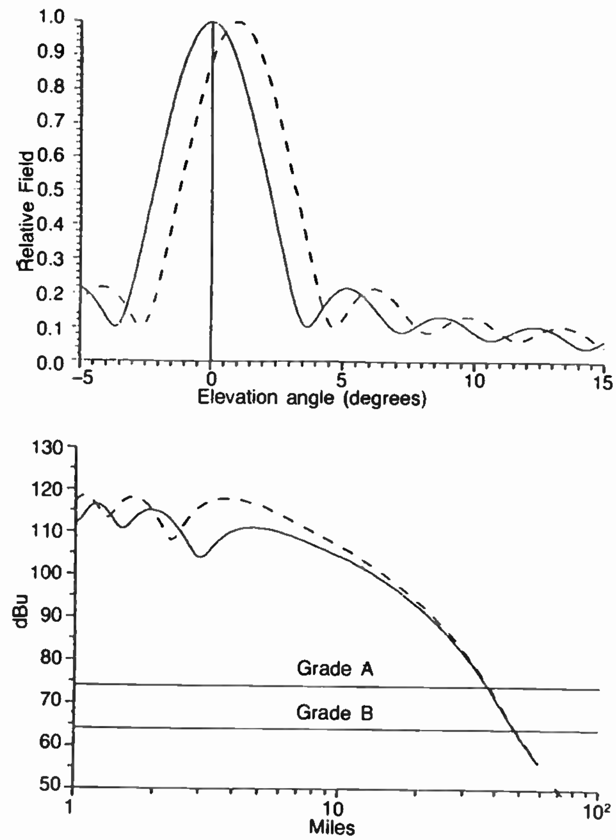
Testing the antenna system determines that:

- The transmission line and components are properly assembled
- The reflections from the antenna and other components at or near the tower top are sufficiently low so that no visible ghost occurs



Curve	Null Fill	Transmitting Power	Antenna Gain	ERP
....	5%	120 kW	20	2,400 kW
— — —	10%	120 kW	20	2,400 kW
———	20%	120 kW	20	2,400 kW

Figure 6.8-17. Null fill.



Curve	Beam Tilt	Null Fill	Transmitting Power	Antenna Gain	ERP
— — —	1	10%	120 kW	20	2,400 kW
———	0	10%	120 kW	20	2,400 kW

Figure 6.8-18. Null fill and beam tilt.

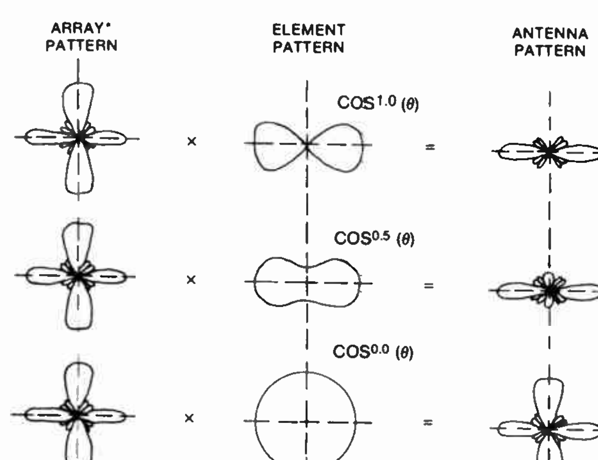
- The impedance presented to the transmitter will result in the maximum transfer of energy.

For an extremely broadband device, like a coaxial transmission line which is usually designed to cover the entire, or at least, a large portion of the TV band, the time-domain reflection (TDR) is the most effective test to determine if the line and components have been properly assembled.

Note that VSWR measurements made at the input to the transmission line measure the combined input impedance at that point only, and not the discreet reflections that can cause ghosts.

The TDR, on the other hand, will measure the reflection on a time rather than frequency basis so the exact position and magnitude of the reflection can be determined.

A high VSWR at this point is an indication of a mismatch between transmission line and antenna. For example, a VSWR of 1.22 is a return loss of 20 dB



*Six Bays, One Wavelength Spacing

Figure 6.8-19. Pattern multiplication.

Table 6.8-2
Distance to 1 mW/cm² level.

Service	Low V	High V	UHF
ERP	100 kW	316 kW	5,000 kW
EIRP	164 kW	518 kW	8,200 kW
Distance to Nearby Tower	89 ft	163 ft	624 ft
Minimum Height Above Ground	43 ft	72 ft	305 ft

Table 6.8-3
Reflection Comparison.

Location of Reflection	Lower Elbow Complex	Antenna
Distance	100 ft	1,000 ft
VSWR	1.08	1.04
Reflection r	4%	2%
Δf MHz	4.96	0.49
Displacement in μsec on TV Receiver	0.2	2.0

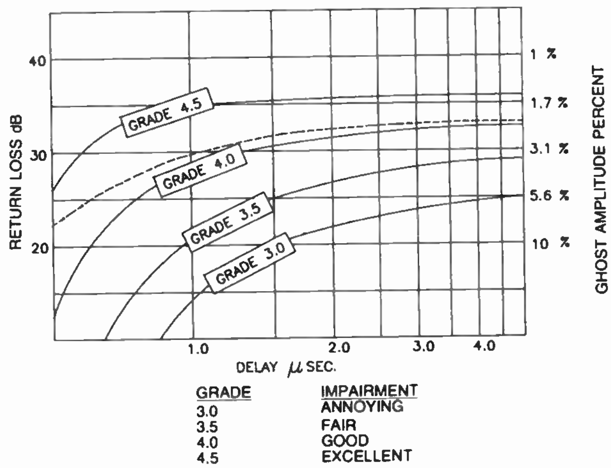


Figure 6.8-20. Signal impairment grades.

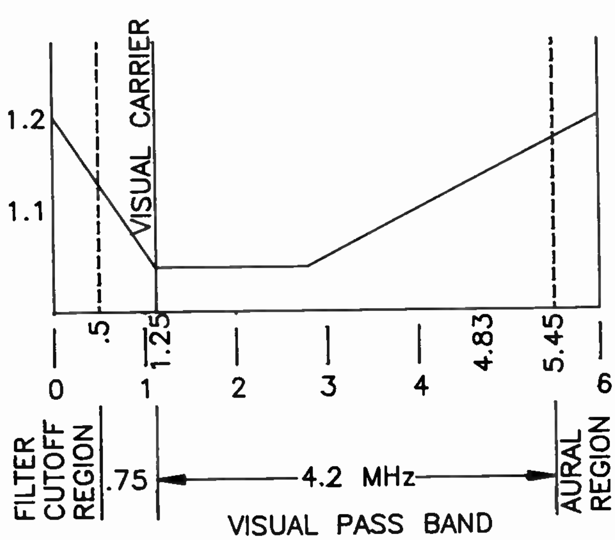


Figure 6.8-21. Acceptable VSWR levels.

(99% of the power is transferred to the antenna system and 1% is reflected).

There are usually two sources of reflection assuming that the line itself is reflectionless.

- Lower elbow complex—VI
- Upper elbow complex and antennas—V2

When measured at the input to the system these can be thought of as two vectors. Each vector individually will rotate about the preceding vector at a frequency corresponding to the distance from the point of measurement (see Figure 6.8-22).

The frequency of rotation can be calculated as follows:

$$\Delta f = \frac{496^*}{L} \text{ coax}$$

$$\Delta f = \frac{496^*}{L} (\lambda_o/\lambda_g) \text{ waveguide}$$

*Half the velocity of light. L in feet.

Close reflections (approximately 100 ft) will have a period equivalent to 5 MHz; distant reflections (approximately 1,000 ft) will have a period equivalent to 0.5 MHz. The individual and combined reflections are shown in Figure 6.8-23.

The use of a *network analyzer* with TDR is effective in determining the location and magnitude of the antenna system VSWR and sources of reflection. Measurement of VSWR in both frequency and time domains for an antenna and 1,700 ft of transmission line (waveguide) is shown in Figure 6.8-24 (a) and (b). The near and far reflections are quite apparent in the frequency domain.

At UHF frequencies where the transmission line guide wavelength is 2 to 3 ft a 50°C temperature change on a 1,000 ft tower can result in electrical length changes of λ/4 to λ/2 wavelength. This is significant in

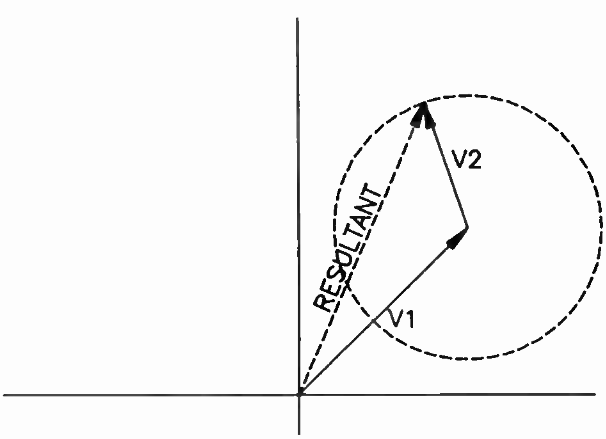


Figure 6.8-22. Vector relationship.

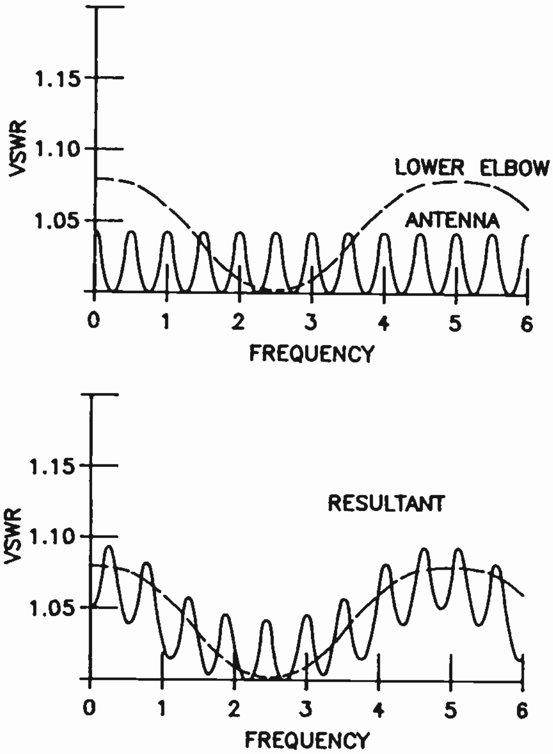


Figure 6.8-23. Example of combined reflections.

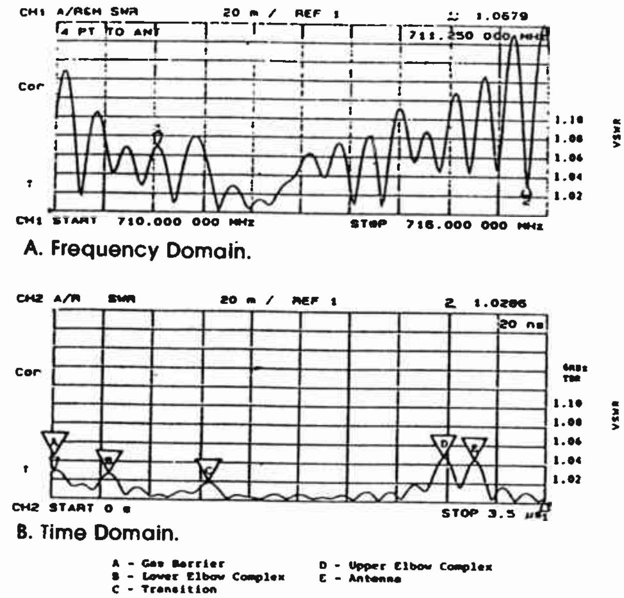


Figure 6.8-24. VSWR of antenna and waveguide (1,700 ft) frequency domain and time domains.

dual feed systems where the differential change can cause beam tilt.

Deflection and Wind Load

Guy tension in guyed towers is usually adjusted so that the tower deflects as a straight member.

Towers for broadcast service, when so specified, are designed for maximum deflection of 0.5°, which means that the top plate will deflect this amount for the maximum wind velocity. For instance, a 40 lb tower will thus deflect 0.5° for a 100 mph wind. Since tower deflection varies as the square of the wind velocity, the deflection will be 0.125° for a 50 mph wind.

Structurally a free standing antenna can be considered as a cantilever beam in which the deflection increases toward the end. Antenna deflection is stated as the angle from the vertical of the chord that connects the base to the top of the antenna.

The movement of the antenna beam and the variation in signal strength for a 50 mph and 100 mph wind is shown in Figure 6.8-25 and summarized in Figure 6.8-26.

The 50 mph wind condition is one that may occur 25 times a year at a 1,000 ft elevation, and about four times a year at a 500 ft elevation (see Table 6.8-4).

The 100 mph wind is a design limit figure which rarely occurs, and one during which there would probably be little television viewing. Most outdoor receiving antennas would probably be severely damaged in such a wind, and power service seriously curtailed.

Antenna Feed Systems

The feed system of a television broadcast antenna is that portion of the transmission system having its input at the antenna terminal (which is at the top of the vertical run of coaxial transmission line in the tower) and its output at the radiating elements.

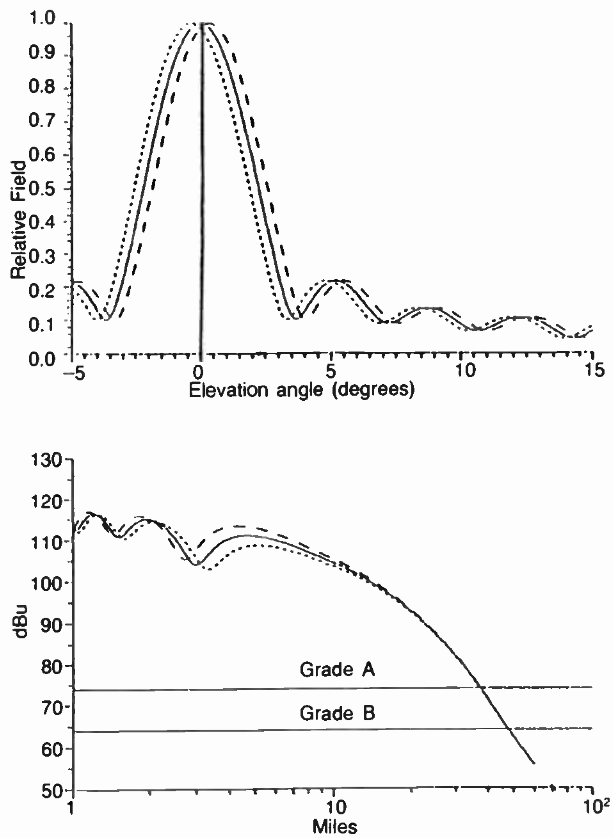
Most antenna gain specified by the manufacturer takes the losses of the feed system into account. Therefore, when system gains are calculated, the feed system loss can be excluded.

In the television broadcasting field, three types of feed systems are in wide use: *branching*, *standing wave* and *traveling wave feed systems*.

Branching Feed Systems

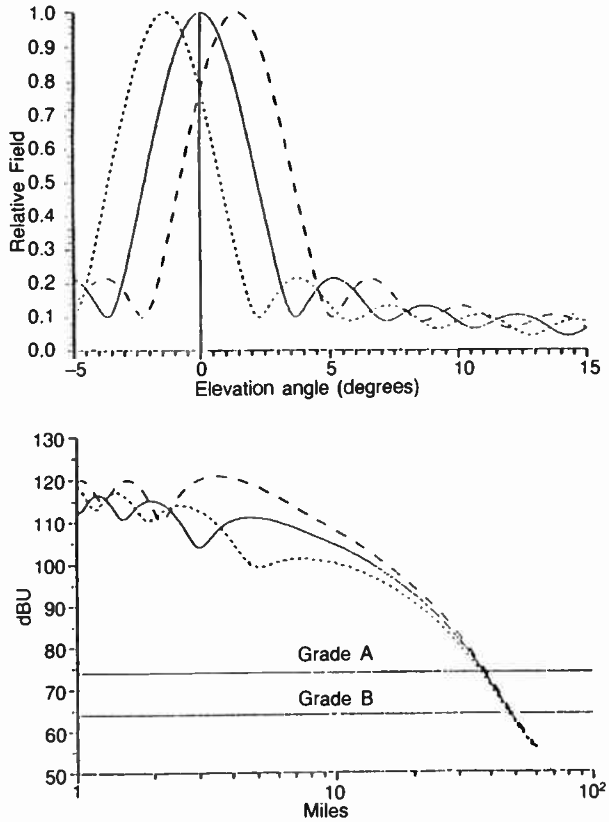
Branching feed systems are necessary for all-band antennas (an antenna that covers the full VHF or UHF band) and since they are center fed, they eliminate beam steering or beam tilt.

This feed system progressively divides the power as shown in Figure 6.8-27. It is used when the radiators are individual elements, each with their own terminal—such as a dipole or panel. The system shown in (Figure 6.8-27(a)) will have a narrower impedance bandwidth than (Figure 6.8-27(b)) since, for economy, one eight-way power divider is used. The junction impedance is $Z_i/8$. If the element impedance is 50 Ω, the power divider must transform 6.25 Ω to 50 Ω. The system shown in (Figure 6.8-27(b)) uses two-way power dividers and is sometimes called a *corporate feed*. Although it has a broader bandwidth, it is less economical since there are seven power dividers with additional interconnecting cables. Null fill and beam



Curve	Wind	Transmitter Power	Antenna Gain	ERP
....	Away	120 kW	20	2,400 kW
— — —	Forward	120 kW	20	2,400 kW
————	0	120 kW	20	2,400 kW

Figure 6.8-25. Field strength changes with a 50 mph wind.



Curve	Wind	Transmitter Power	Antenna Gain	ERP
....	Away	120 kW	20	2,400 kW
— — —	Forward	120 kW	20	2,400 kW
————	0	120 kW	20	2,400 kW

Figure 6.8-26. Field strength changes with 100 mph wind.

tilt is accomplished by changing the length of the feed cables or using unequal power dividers, or both.

A problem branching feed systems is the presence of the feed line in the aperture of the lower elements. The feed lines can cause reradiation and distort the azimuth pattern. The branching feed system can be more effectively used with panel antennas that require a center support tower or mast where the transmission lines are behind the antenna.

Standing Wave Feed Systems

A coaxial or waveguide transmission line can be shorted at the far end, resulting in standing waves along the length of the line. If slots or coupling probes are appropriately sized and positioned, the RF energy can be radiated and a desired amplitude and phase distribution across the aperture can be obtained. This resonant array structure has a desirable feature; all coupling parameters are the same and equally spaced. Its disadvantages are a narrow bandwidth, and it can only be used at high VHF and UHF frequencies (Figure 6.8-28).

Traveling Wave Feed

The traveling wave feed system operates on the principle of a gradual attenuation (radiation) of the input signal as it progresses from the input along the aperture of the antenna. An application of this principle is the *slot antenna* or *spiral antenna* (see Figure 6.8-29).

Table 6.8-4
Antenna and tower deflection.

Wind Velocity (mph)	50	100
Wind Load (psf)	10	40
Deflection (in) Antenna at Top	4.9	19.6
Deflection (degrees)		
Chord Bottom to Top	0.227	0.914
Tower	0.125	0.500
Antenna and Tower	0.352	1.414

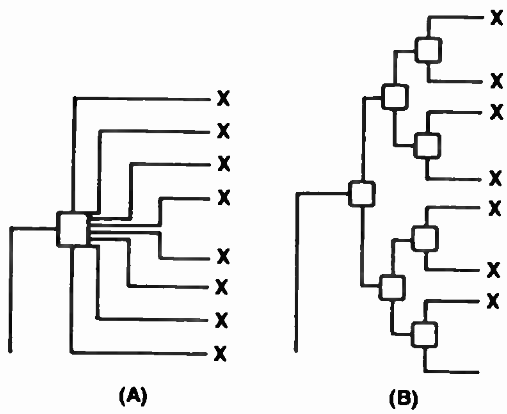


Figure 6.8-27. Branching feed systems.

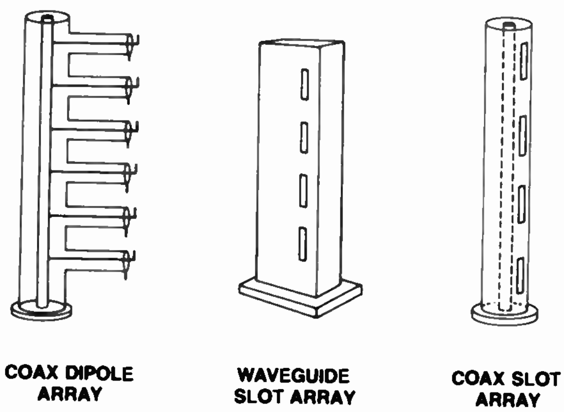


Figure 6.8-28. Slotted arrays.

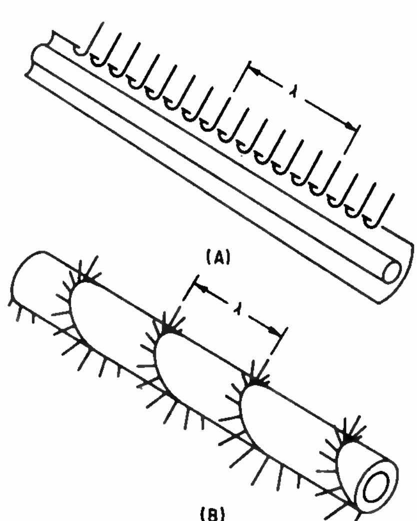


Figure 6.8-29. Traveling wave excitation.

Figure 6.8-29(a) shows the principle of this feed system using short rod radiators to illustrate the theory. A number of uniformly spaced radiators per wavelength are loosely coupled to a coaxial line. Because of the number of radiators and the relatively slight reflection between them, the effect is essentially that of a uniform loading. The result is a uniformly attenuated traveling wave in the line. Since a traveling wave has a linear phase characteristic, the excitation of each successive radiator will be lagging from the previous one by an amount, which depends on the spacing between the radiators and the velocity of propagation in the line. If the radiators are alike, their currents will have the same phase relationship as the excitation. Thus the radiating currents will be successively lagging, and repetition of phase occurs after every guide wavelength.

To obtain an omnidirectional pattern, the radiators, instead of being in line, can be moved around the periphery to form a *spiral* as shown in Figure 6.8-29(b). For a horizontal main beam, the pitch of the spiral has to be equal to the guide wavelength in the transmission line. In this arrangement all the radiators in any one vertical plane on one side are in phase, and the phase difference between radiators in different planes equals the azimuth angle difference between the planes. That is, the phase rotates around the periphery. The rotating phase produces a rotating field. Because of the relatively small amount of current change from layer to layer, an onmidirectional pattern is produced.

ANTENNA TYPES

The types of antennas available can be grouped into the following categories:

- Top or side mounted
- Linear or circular polarization
- High power, low power

Antenna types and characteristics are summarized in Table 6.8-5.

The prime purpose of the tower is to support the antenna. The antenna should have the required radiation characteristics to deliver a satisfactory signal to the viewer. Since it is desired to cover as large a viewing area as possible, tall towers are used. It is not uncommon for antennas to be mounted on 1,000 ft, 1,500 ft or 2,000 ft towers. These tall towers are guyed triangular towers with face dimensions of 5 to 10 ft.

The cost of a tower is heavily dependent on the wind load presented by the tower the transmission line and antenna. For VHF antennas, the antenna wind load is the most significant parameter. For UHF antennas the transmission line or waveguide is the more significant factor.

The length of the antenna is related to the channel, or wavelength and the gain requirements (Table 6.8-6).

The ideal onmidirectional antenna would be a small diameter, infinitely stiff pole. The smallest cross sec-

Table 6.8-5
Commonly used television transmitting antennas.

	Horizontal Polarization		Circular Polarization	
	Top Mtd	Side or Tower Mtd	Top Mtd	Side or Tower Mtd
VHF Low Band	Batwing	Butterfly Doublett H-Panel	TDM	
VHF High Band	Batwing Traveling Wave	Butterfly Doublett H-Panel	Spiral Slot&Dipole Slot&Director	TCP CBR Arrowhead Ring
UHF	Slot Coax Wave Star Trasar	Doublett	Slot&Z Dipole Slot&Dipole Slot&Director	Slot&Dipole Slot&Dipole Slot&Director

tion that can resonant is $\lambda/2$. Therefore, the antenna cross-section is also dependent on the wavelength.

At low VHF frequencies the resonant half wave element of 8 ft is sufficiently large so that a support pole can be used in the center with the antenna mounted outside (such as a batwing antenna). On the other hand, at UHF frequencies the resonant half wavelength of 1 ft is so small that the resonant elements must be outside the support pole or the support pole itself must be a radiator like a slotted coax or waveguide array.

Ideally, omnidirectional antennas should be fed from the bottom, so there are no feed lines in the aperture of the antenna to distort the pattern. If feed lines are required for the upper elements, they should be on the inside. Of course, this is not possible for waveguide arrays. A conflicting requirement is that the diameter of the feed line be large enough to satisfy the power handling requirements, yet small enough not to create pattern distortion.

To minimize pattern distortion by the tower, the antennas should be mounted on top of the tower rather than on the side

The wavelength of high VHF and UHF antennas is such that the resonant half wavelength is too small to be a satisfactory structural member.

The radiating element for many UHF antennas is either the slot or dipole excited from an internal source such as coaxial or waveguide.

Most slotted arrays are traveling wave structures which are bottom fed and contain many slots. The greater the number of slots the narrower the bandwidth, and the greater the beam steering or tilt between the visual and aural frequencies.

Some slotted arrays use the outer conductor (coax), or the waveguide itself as the structural member.

Table 6.8-6
Dimensions of typical antenna.

	Wavelength	Length for Maximum Gain	Maximum Gain
Low VHF	16 ft	100 ft	6
High VHF	5 ft	90 ft	18
UHF	2 ft	120 ft	60

Panel antennas are fed from the back of the panel and require a secondary structural member to support the panels. This can be the tower itself. Three or four panels must be mounted around the tower to produce an omnidirectional pattern.

Panel antennas, when fed with a corporate type feed harness, can preserve the wide bandwidth of the dipole and cover the full high VHF band (174–216 MHz) or UHF band (470–800 MHz).

VHF ANTENNAS

Linear Polarized

Superturnstile/Batwing

The first antenna developed for commercial service was the *Superturnstile*. It consists of a central sectionalized steel pole upon which is mounted the individual radiators, or *batwings*. These radiators are mounted in groups of four around the pole in north-south and east-west planes to form a *section*. The sections are stacked one above the other to obtain the desired gain. Figure 6.8-30 illustrates this construction.

Each of the radiators at the batwing antenna is fed by its own feed line. The impedances are carefully

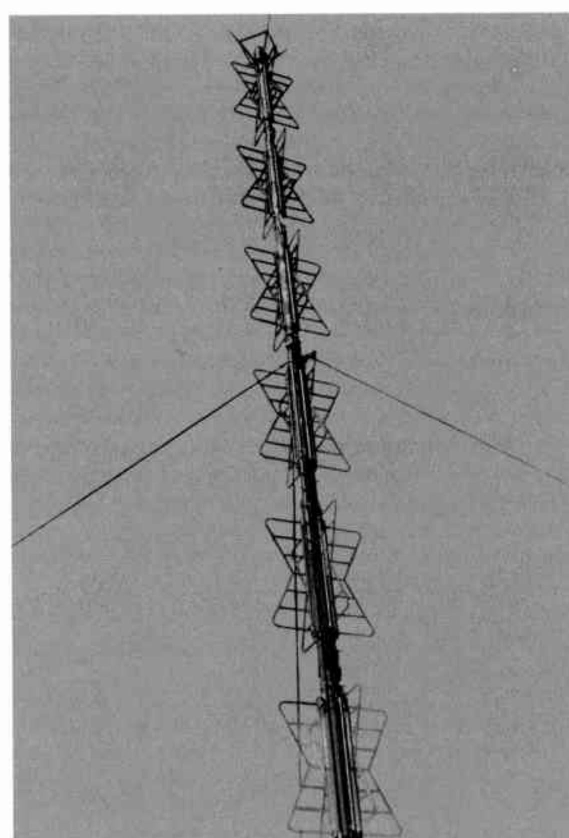


Figure 6.8-30. Superturnstile/batwing antenna (Courtesy of Dielectric).

matched. The feed lines, in turn, are combined at junction boxes which perform the dual function of feeding power simultaneously to all feed lines, and transforming the combined impedance of these lines to that of the transmission line. This latter function is achieved by the use of three-stage transformers immediately below the junction box.

At the base of the antenna, a combining network is used when there are more than two junction boxes. These networks accomplish power division between portions of the antenna, if desired. Batwing antennas are manufactured in various gains from 3 to 12 for channels 2 to 6 and gains of 6 to 18 for channels 7 to 13. They can also be designed for various types of null fill and they have been used in stacked and candelabra installations.

The batwing antenna is wideband and can be used for two channels. A number of them are operating at channels 4 and 5 or channel 6, and an FM channel and also in various combinations in the channel 7 to 13 range.

Traveling Wave

The traveling wave (TW) antenna is a slot antenna with a traveling wave feeding the slots (see Figure 6.8-31).

The TW antenna is a coaxial line, with pairs of slots in the outer conductor spaced at intervals of a quarter wavelength throughout its length. Probes at the center of each slot distort the field within the line to place voltages across the slots. These, in turn, drive currents on the periphery, setting up a radiated field. Attenuation of the signal by withdrawal of a portion of the power at each slot, reduces it to a very low value at the upper end of the antenna. There, a special pair of slots, designed to match the line, radiates the remaining power.

Operation of the TW antenna can be better understood if the section of the aperture having pairs of slots are recognized as being, in effect, dipoles.

Successive pairs of slots are alternately in one plane and in another at 90° to it, so that the antenna can be simulated by stacked dipoles with a 90° angle between successive layers.

In a given plane, reversal of the direction of feed every half wavelength (by placing the probes on opposite sides of the slots) together with the half wave change in phase of the signal as it passes along the aperture through this distance, results in all the *dipoles* in that plane being fed in phase.

The same action takes place in the other plane, except that they are fed 90° out of phase with the first plane, owing to their 90° displacement along the antenna.

Each plane of dipoles radiates essentially a figure eight pattern. Since the planes are fed in quadrature, addition of the patterns results in an omnidirectional pattern. Because of the circular cross-section and the lack of obstructing radiators, the resulting horizontal pattern is almost nearly circular.

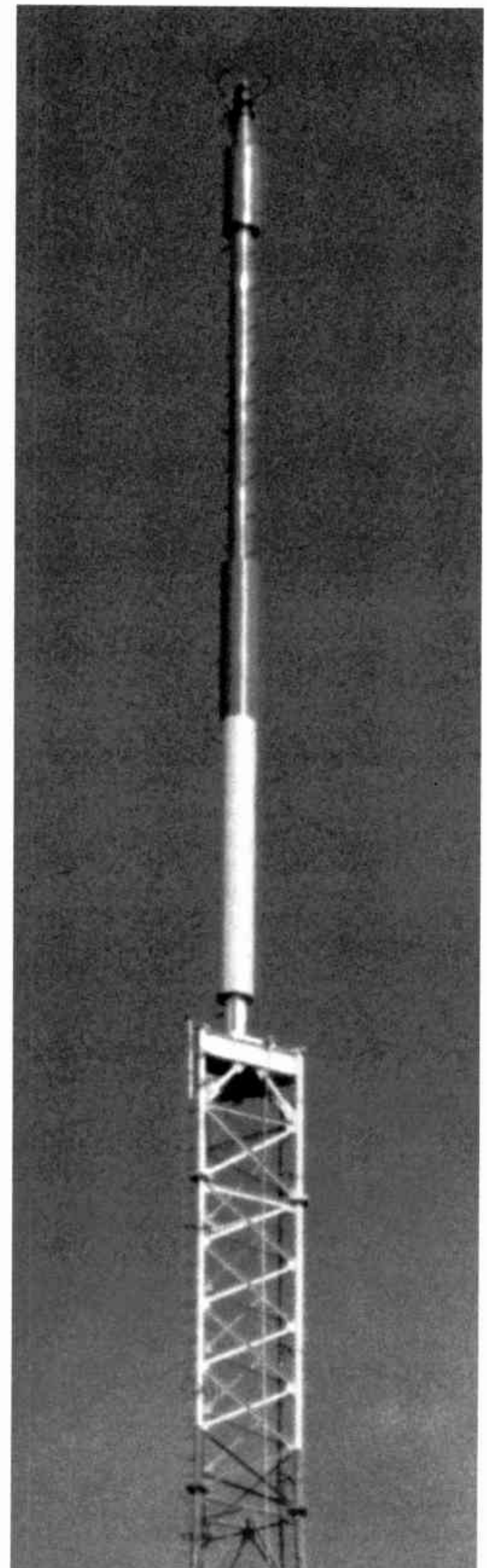


Figure 6.8-31. Traveling wave antenna (Courtesy Dielectric).

Panels

The panel antenna is designed to wrap around the tower. Four panels are needed for a square tower where each panel must radiate a $\cos^2(\theta)$ element pattern. For a triangular tower, each element must radiate a $\cos(\theta)$ element pattern for an omnidirectional pattern.

A wide variety of azimuth and elevation patterns can be obtained by using fewer panels or changing the power division to the panels.

The panels are 0.7λ to 0.9λ in vertical length and spaced approximately one wavelength.

The radiating elements may be either single dipoles like the H-panel *rhombus* in Figure 6.8-32(a) or *delta dipole* (*butterfly*) in Figure 6.8-32(b) which is essentially a folded back batwing or dual dipoles.

The *dual dipole* in Figure 6.8-32(c) is designed to minimize downward radiation and consists of a pair of dipoles spaced a half wavelength apart.

The impedance bandwidth at the panel antenna is very wide and capable of handling more than one TV channel. The bandwidth can be further improved with the four-side configuration by using phase rotation. That is, 90° between each panel in the bay. The impedance at the four-way power divider will be conjugate for every two panels, resulting in wider bay impedance bandwidth than the element itself.

Circular Polarized

Transmission Dual Mode (TDM)

The TDM shown in Figure 6.8-33 is a circularly polarized antenna for channel 2 to 6, and is designed for tower top mounting. It is capable of replacing an existing six-bay Superturnstile without any increase in tower windloading.

The TDM antenna utilizes seven layers of radiators in a slanted dipole configuration with three radiators mounted symmetrically around the pole per layer. Each of the three radiators is fed in phase by a single feedline. Only 21 feedlines are required for the entire antenna. A branch type feed system is used to achieve excellent vertical pattern stability. One junction box feeds the upper four layers and another feeds the lower three layers, each box fed by a 3 1/8" line. This feature allows for standby capability in the event of weather related or other damage. The TDM can be supplied with deicers and/or radomes depending upon environmental requirements.

The unique design of the TDM provides for excellent pattern circularity and axial ratio. Axial ratio measurements are performed on a complete full scale as-built antenna standing vertically at the factory. Since all elements are excited, the mutual effects of adjacent elements are considered.

Circularly-Polarized "V" (CPV)

The CPV antenna shown in Figure 6.8-34 is a circular, polarized top-mounted antenna consisting of three cross "V" dipoles mounted at 120° intervals around a vertical mast. The dipoles are segmented by three vertical grids like a corner reflector, used both for isolation and to shape the element for good circularity. The cross dipoles are fed in phase quadrature and radiate circular polarization from each element.

A branching feed system is used with the lines fed up the mast. Null fill and beam tilt are accomplished by changing the electrical length of the feed cables.

Panel

The cross dipole panel antenna consists of dual dipole feed in both space and time quadrature, a necessary condition for circular polarization.

The transmission circularly polarized (TCP) series uses dipoles shown in Figure 6.8-35(a) and are in the form of a clover leaf mounted on front of a ground plane or panel. The four elements generate the required $\cos(\theta)$ element pattern. By tilting the ground plane edges forward, as a partial corner reflector, further shaping can be obtained.

The circularly basket reflector (CBR) series in Figure 6.8-35(b) consists of fat dipoles mounted in a basket. The element pattern is controlled by the diameter and depth of basket. The ring panel antenna in Figure 6.8-35(c) is a large resonant loop mounted against a ground plane. The ring is approximately one wavelength circumference and hence radiate a circular polarized wave in all planes perpendicular to the panel.

The arrowhead dipole in Figure 6.8-35(d) is similar to the other cross dipoles. The shaping of the element pattern is accomplished by tilting back the dipole to broaden the E-plane pattern.

Spiral

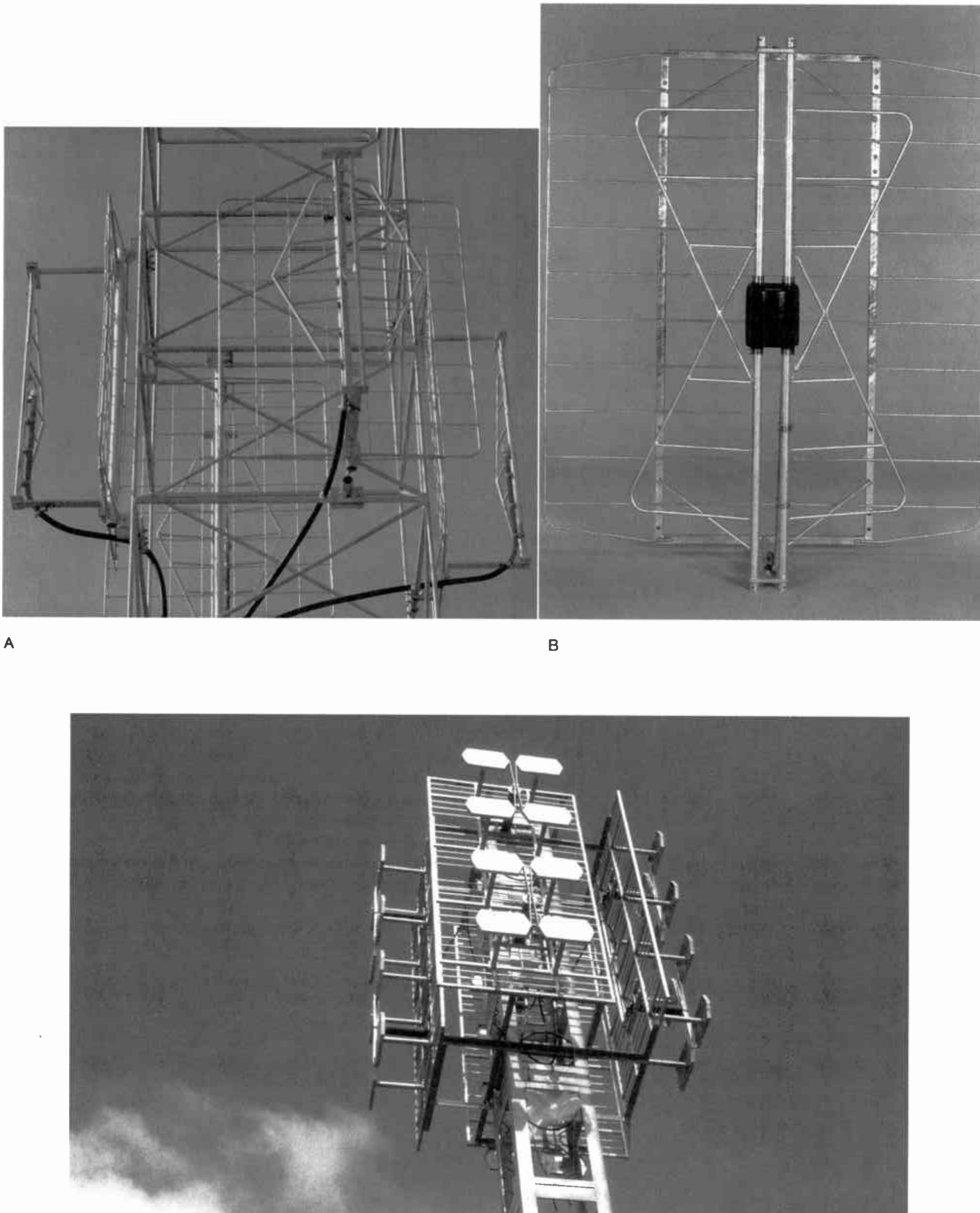
The spiral antenna is designed for tower top mounting. The antenna is composed of sections, each operating under the principals of a traveling wave antenna. Each section consists of two or more stainless steel coils (radiating elements) wound around the supporting pole. A *beltline* feed system is used in each section to excite the radiating elements. The phase initiating the traveling wave is a function of the number of coils used.

Each section is terminated with radiating end loads on each coil to radiate the remaining energy. These loads minimize the reflections of energy from the far end of the radiator back toward the input, which would affect the traveling wave illumination and distort the pattern.

Vertical pattern characteristics such as gain, null fill and beam tilt are determined by a number of factors; the number of wavelengths *wrapped* around the pole, the spacing of the radiators off the pole, the pitch (angle of wrap) around the pole and the phasing between the vertical sections. The antenna is deiced by low voltage, high current dc to heat the radiating coils. A spiral antenna under test is shown in Figure 6.8-36.

Slotted Coax

The circular polarized slotted antenna in Figure 6.8-37 is similar to the linear polarized slotted antenna with the addition of vertical radiators.



A

B

C

Figure 6.8-32. Panel antennas (a) H-panel Rhombus (Dielectric), (b) Delta dipole "Butterfly" (Harris), (c) dual dipole (MCI).

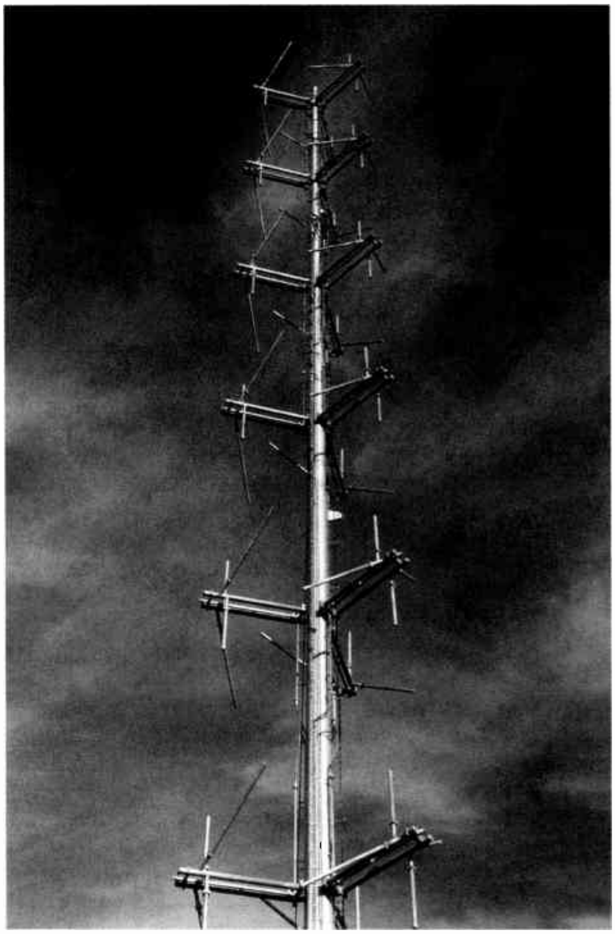


Figure 6.8-33. TDM antenna (Courtesy Dielectric).

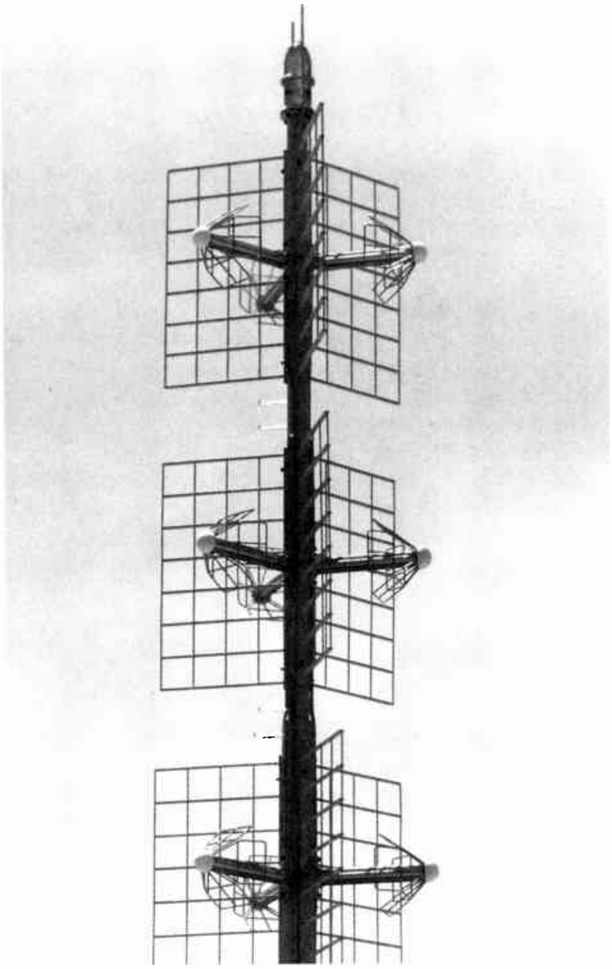


Figure 6.8-34. CPV antenna (Courtesy Harris).

The vertical radiators are located outside the array and the energy coupled to them from the slot is in phase quadrature.

Any polarization from elliptical to circular can be obtained by adjusting the length of the radiators.

UHF ANTENNAS

Linear Polarized

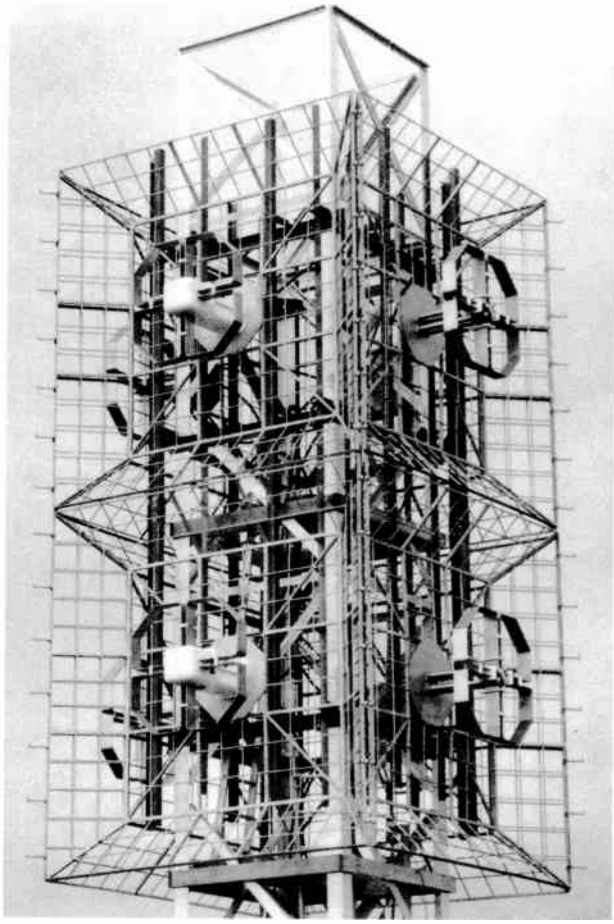
Slotted Coaxial

The UHF *pylon antenna* in Figure 6.8-38 is a coaxial transmission line with radiating slots in the outer conductor. The number of slots (per layer) around the circumference is determined by the horizontal pattern; such as one slot for a skull shaped pattern, two for a peanut shaped pattern, three for a *trilobe* pattern and four or more slots, depending on outer cylinder diameter, for an omnidirectional pattern.

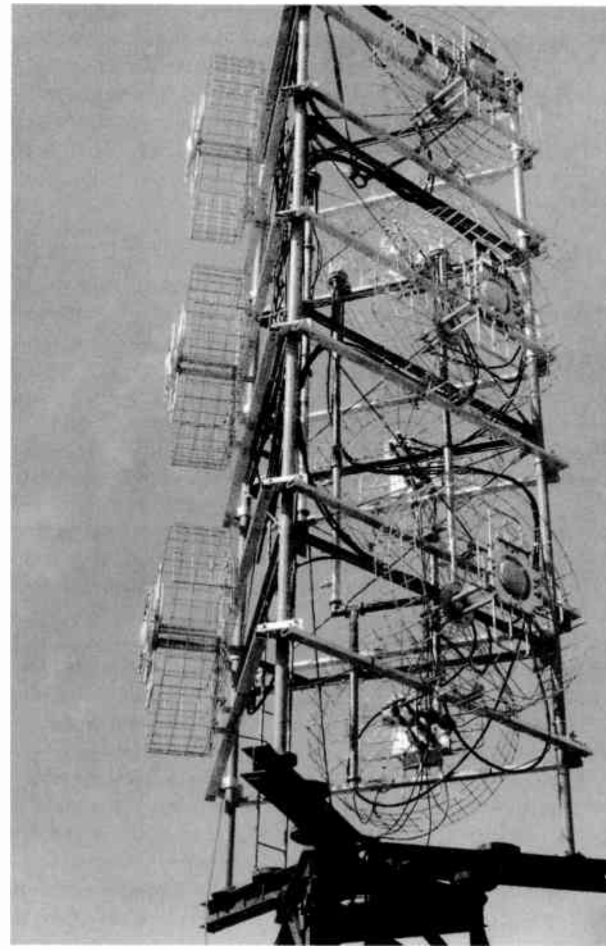
The layers are located at one wavelength spacings along the antenna, with the number of layers determined by the vertical gain and pattern. The radiation parameters of phase and amplitude are determined by a combination of slot length and coupler bar diameter. This allows discreet control of the illumination along the antenna aperture at every wavelength resulting in vertical pattern control and shaping. It also allows for maximum aperture efficiency and, in conjunction with the extremely low cross-polarized radiation component of a slot, produces the highest vertical gain for a given antenna length. The antenna is a bottom fed traveling wave resonant antenna. Some antennas *launch* the energy into the coax radiating section at the center. Others *feed* the coax radiating section at the center. The bottom coax feed is located inside the radiating coax feed. The pylon uses a radome to cover the radiating slots only.

One version is omnidirectional bottom fed traveling wave waveguide slot antenna. Since it is waveguide, no inner conductor is required. The signal propagates in the TM mode. The resultant current rings on the inside wall are interrupted by slots cut in the wall. The slots have radome covers.

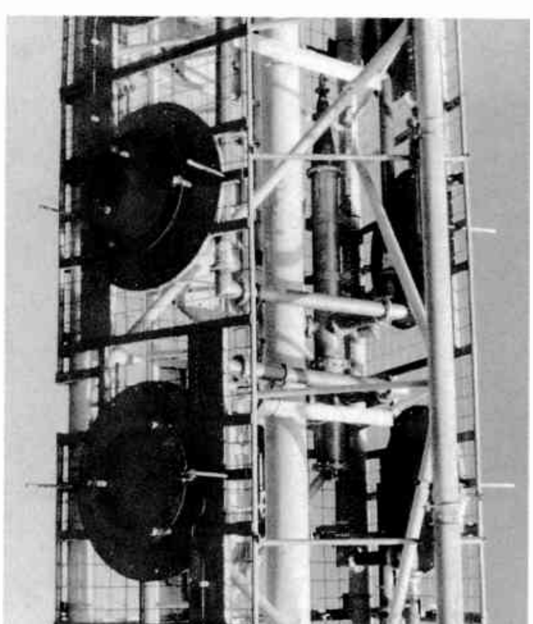
Another version is designed to radiate a cardioid pattern and has a single row of slots. This is also a bottom fed traveling wave waveguide array operating in the TE_{01} mode. This is built by exciting the fields



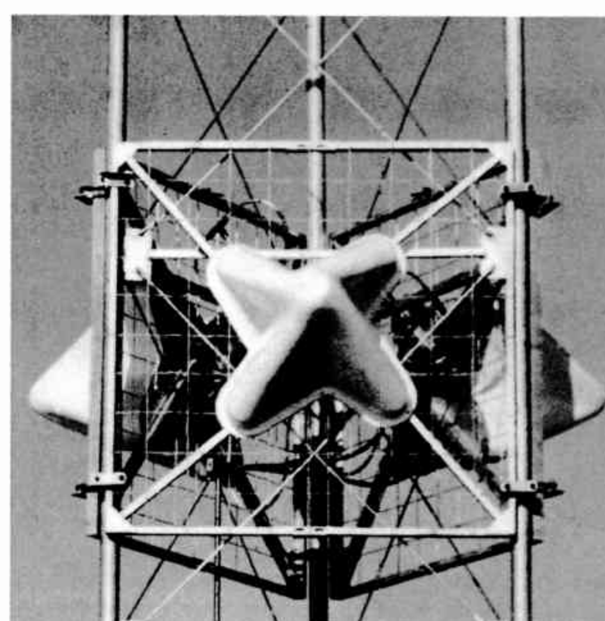
A



B



C



D

Figure 6.8-35. Panel antennas (a) TCP (Dielectric), (b) CBR (Harris), (c) ring panel (Lampro), (d) arrowhead dipole (MCI).

Figure 6.8-36. Spiral antenna on test range (Courtesy Dielectric).

in a rectangular waveguide and rolling the waveguide into a cylinder.

Slotted Director

The *slot director antenna* in Figure 6.8-39 is linearly polarized with a single row of slots with directors to fill in the low signal on the opposite side of the slots.

The slots are one wavelength long on one wavelength centers. Each is fed directly at the slot from a high-power strip line feed. Each section is up to 8λ's

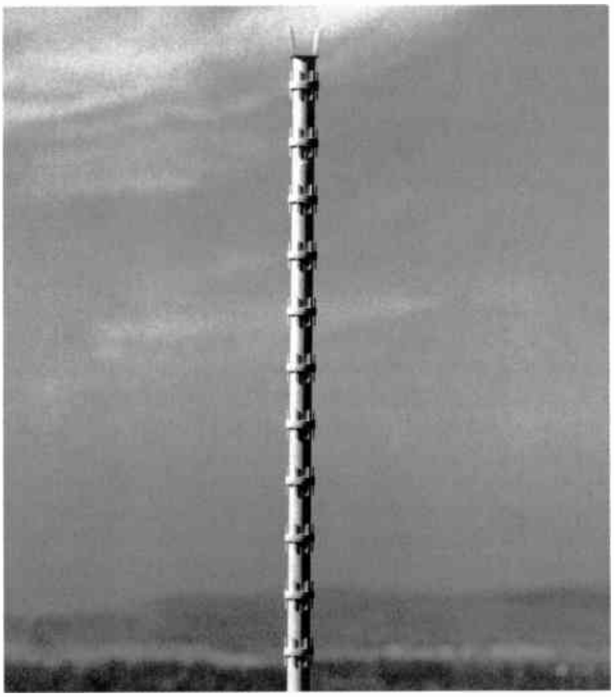


Figure 6.8-37. CP slotted antenna (Courtesy Andrew).

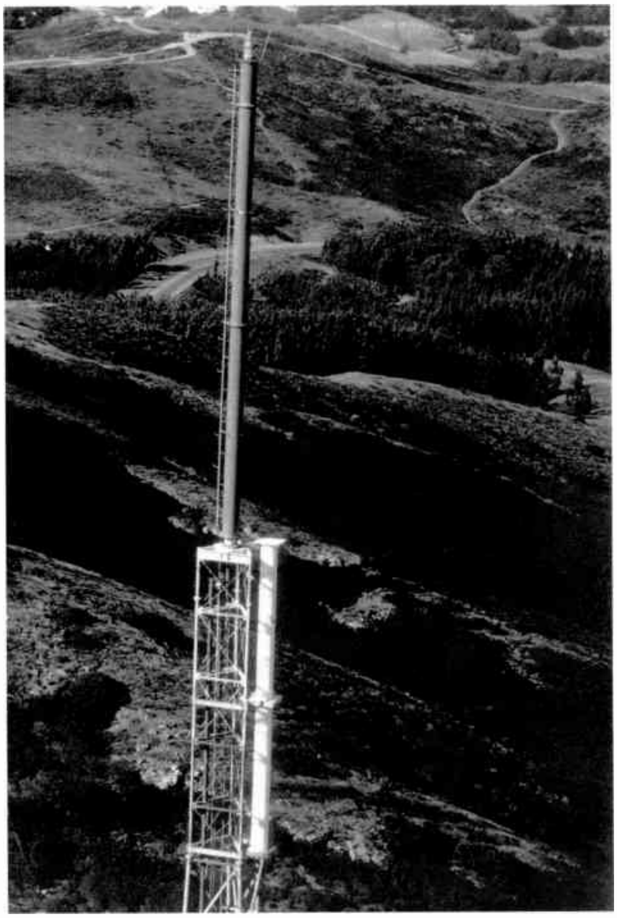


Figure 6.8-38. UHF Pylon antenna (Courtesy Andrew).

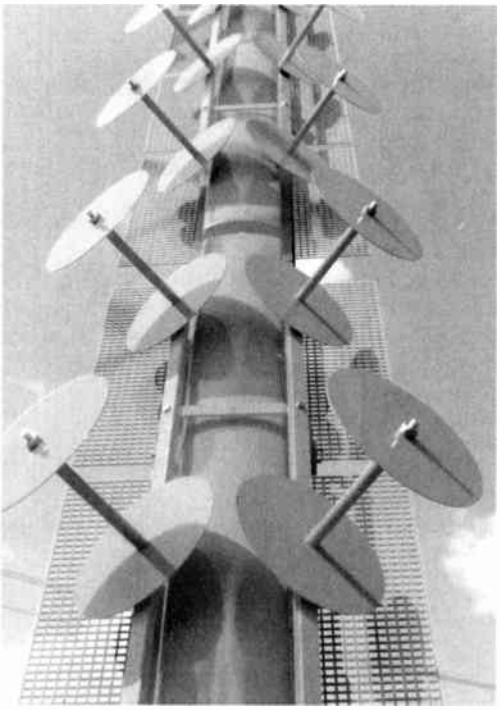


Figure 6.8-39. Slotted director antenna (Courtesy Cablewave).

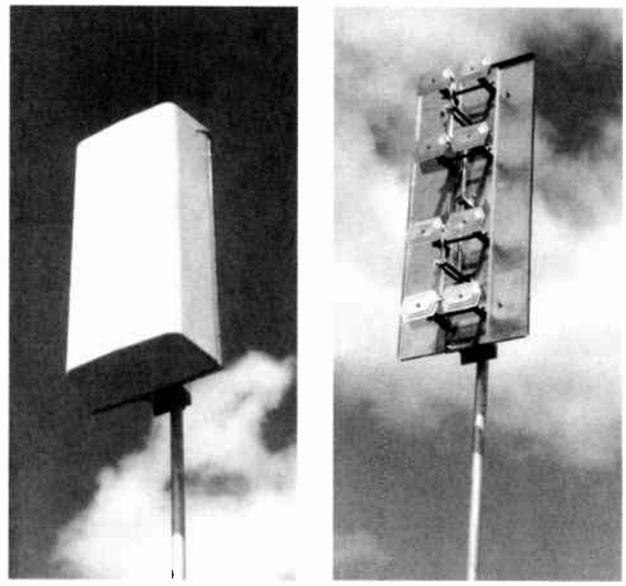


Figure 6.8-40. Doublet array antenna (Courtesy MCI).

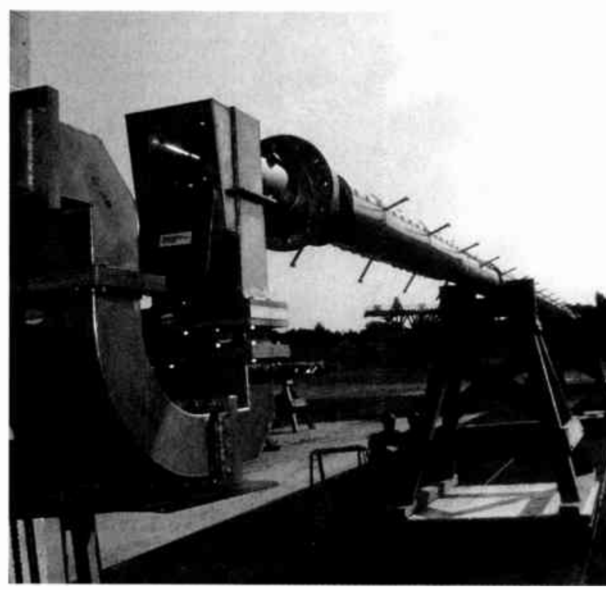


Figure 6.8-41. E-type antenna.

and four sections are usually stacked for the desired gain.

Each section is branch fed while the slots in each section are fed in a traveling wave manner.

Doublet Array

The doublet array panel antenna in Figure 6.8-40 consists of two pairs of dual dipoles, each pair spaced slightly greater than one wavelength apart.

The dual dipole antenna is two broadband dipoles spaced half wavelength apart.

The half wavelength spacing reduces the coupling and cancels the downward radiation thereby increasing the gain.

The doublet array has a VSWR less than 1.10:1 from 470 to 800 MHz.

Arrays composed of doublet arrays can be arranged to produce omni or directional patterns, and can handle four or more high power channels.

E-Type Antenna

The E-type antenna is a high power version of the slotted coaxial antenna (see Figure 6.8-41). It uses a bottom, end feed design incorporating a high power waveguide to coax transition. This specially designed transition allows the antenna to accommodate powers up to 280 kW peak NTSC for omnidirectional service. The E-type is a true traveling wave design which provides a null free elevation pattern.

Circular Polarized Antennas

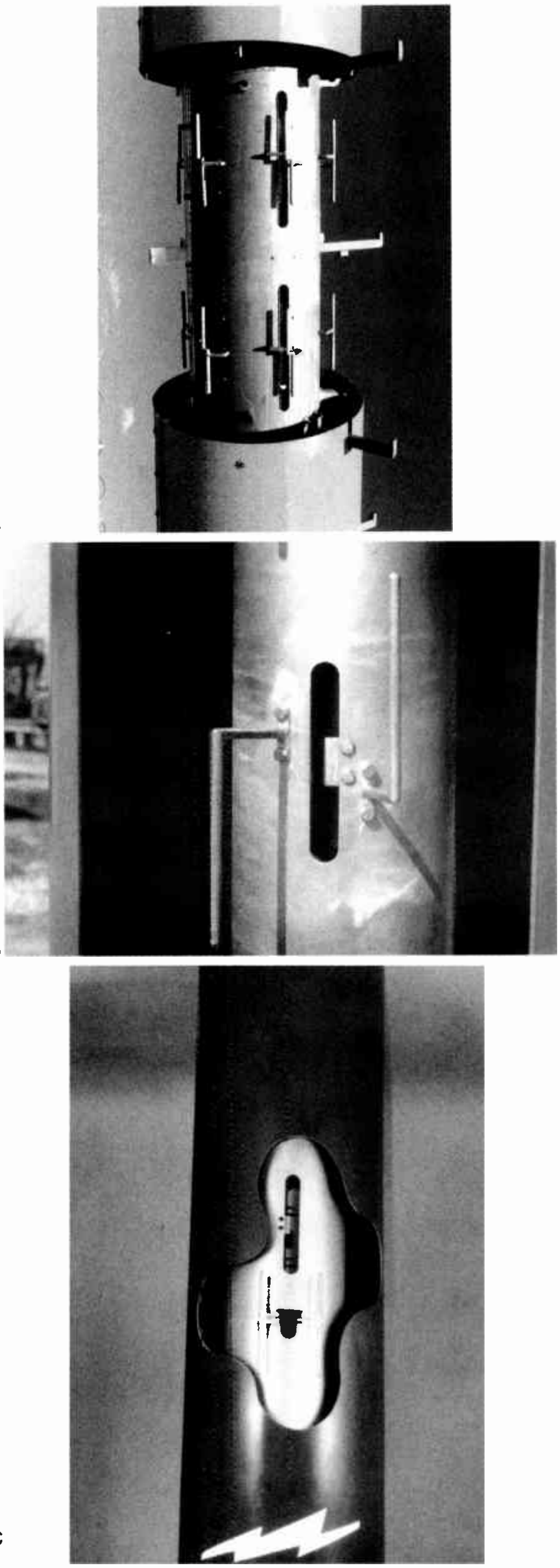
Slotted Coax

The UHF circular polarized slot antennas are similar to the linear polarized slot antenna.

The slot cut in the wall will radiate a horizontally polarized signal. If a vertical dipole is placed above or near the slot, energy will be coupled to the dipole and reradiated as a vertically polarized component.

The number of slots about the periphery of the cylinder can be varied to obtain omni or directional patterns.

The circular polarized pylon antenna in Figure 6.8-42(a) uses a Z dipole located directly above the



A

B

C

Figure 6.8-42. Circular polarized antennas. (a) Z dipole antenna (dielectric), (b) Wave Star vertical dipole antenna (Harris), (c) Trasar vertical dipole antenna (Andrew).

slot, and radiates a vertical component in phase quadrature with the same elevation pattern as the horizontally polarized slot. The size and spacing of the graduated dipole can be used to control the amount of vertical component radiation.

A vertical dipole configuration is shown in Figure 6.8-42(b).

Another configuration consists of a series of slotted arrays that uses a vertical dipole located between the slots, as shown in Figure 6.8-42(c).

The pylon circular polarized type of antennas can be either end fed or center fed. The center fed array can be fed internally for a top mounted location or externally fed for side mounted installations.

The antennas in Figures 6.8-42(b) and 6.8-42(c) are traveling wave circular polarized and each have a TM_{01} mode circular waveguide feed system for very high power applications.

All of the traveling wave slotted arrays have a cylindrical radome covering the full array.

ANTENNA TESTS

Antennas must be tested to determine if the necessary requirements for impedance and patterns are met. Impedance tests are usually run on all production antennas. Pattern tests are normally run on prototype and custom antennas.

Custom antennas are always impedance tested before shipment. These measurements should be made with the antenna completely assembled and in an area free of reflections. Using an impedance plotter or network analyzer a determination of the antenna reflection characteristic can be made.

A pattern measurement test is conducted for two reasons: 1) to determine the gain as compared with a dipole for which a substitution method could be used; and 2) to determine the amount of radiation at all vertical and horizontal angles which have an influence on the coverage.

Pattern tests can conducted on full or partial scale models of the final antenna. Scale models have an advantage of reduced size that permits high gain antennas to be tested in an anechoic chamber, free of reflections (from the ground or nearby objects that occur when conducting full scale tests on a range).

DTV ANTENNA SYSTEM CONSIDERATIONS

The availability of suitable locations for new television transmission towers is diminishing, even in the secondary markets and it is practically nonexistent in major markets. Even after the hurdles of zoning variance and suitable tower location are overcome, FAA restrictions and environmental concerns may delay the construction of a new tower for years. Not surprisingly, many broadcasters are looking at the pros and cons of using existing towers to support their new DTV antennas even though the prime tower top spots are occupied.

For any given antenna, directional or omnidirectional, the tower will modify the as designed antenna patterns. For optimum coverage, the as installed patterns must be known not just at the carrier frequency but throughout the entire channel before the relative position of the antenna and its azimuthal pattern orientation can be fixed. There is usually one position that will provide the optimum coverage without exceeding the structural limitations of the tower. This optimum position can be calculated. Coverage considerations are particularly important to DTV because all undesired energies, such as reflections, translate into a loss of coverage, whereas in NTSC the undesired energies translate primarily into a loss of picture quality.

Another transmission optimization technique that holds promise for DTV is CP. Although mixed results were achieved in the use of CP for NTSC broadcasts, DTV applications may prove more amenable to improvement through the use of this technique. The transmission of CP has obvious drawbacks in the form of a 2× increase in required transmitter power and transmission line, and a more complex antenna. (The 2× increase comes from the fact that ERP is measured in just one radiation plane.) For the DTV signal, *polarization diversity* can be achieved if the vertically polarized signal is transmitted through CP. A polarization diversity system at the antenna can provide missing signal level when one of the horizontal or vertical signal components experiences a deep fade. The inherent diversity attributes of CP operation could, thus, be put to good use in reducing the cliff edge effect of the terrestrial DTV signal.

Implementation Issues

The requirements for DTV antenna type are the same as for NTSC. The antenna should have structural integrity, be capable of mounting on tall towers, be efficient (have maximum gain for minimum length with minimum loss and reflection) and be able to handle power consistent with the ERP and feed line power requirements. In addition the antenna manufacturer must be able to shape the radiation pattern to satisfy coverage and interference requirements.

The FCC had considered using vertical polarization to decrease the coupling between NTSC and DTV antennas. Had this been specified it would have required a new class of antenna to be developed for DTV.

The antennas available today and described in this chapters satisfy all of the DTV requirements including the multi-channel requirements.

NTSC coverage is based on power (ERP) and height (HAAT) using the FCC F (50, 50) field strength prediction curves. From this simple curve, (one for each band) a grade A and B signal level is determined which is defined as the contours. The contour identified the maximum reach of the station signal from which the demographics of the station is determined.

The new DTV prediction requirements specify a different signal level for determining the contour and uses a different statistical model. If the station is to match the DTV and NTSC coverage, it is necessary to first determine a new NTSC contour using the new statistical model and terrain criteria. The NTSC contour using the new procedure will result in a contour that is not the same as the present NTSC contour and the antenna pattern requirement will very likely be different than the original NTSC pattern. For example, if the original antenna was an omnidirectional pattern, the recalculated pattern may not be omnidirectional; in fact, it may result in a pattern that is not physically realizable. To produce a DTV geographical contour that matches the revised NTSC contour in every case will not be possible as the rules are now written in OET Bulletin #69.

Adjacent channel interference is a function of side band splatter and filter requirements. Satisfying the FCC mask will result in high group delays at band edge. Field measurements especially for high power systems will be necessary to determine the effect the high group delay will have on performance and the requirements for the new filter designs.

BIBLIOGRAPHY

Giardina, J., Vaughan, T. and Neuhaus, J., "True APL Levels," IEEE Broadcast Technology Society, Broadcast Symposium, 1984.

Masters, R.W., "The Superturnstile Antenna," Broadcast News, January 1946.

Siukola, Matti, "Predicting Performance of Candelabra Antenna by Mathematical Analysis," *Broadcast News*, October, 1957.

Siukola, Matti, "Evaluation of Circular Polarized TV Antenna Systems," IEEE, BC-24, March 1978.

Stenberg, J., and Pries, W., "Advances in RF System Measurement Techniques," *NAB Engineering Conference Proceedings*, National Association of Broadcasters, Washington, DC, 1989.

Vaughan, T., Windle, J., "Tall Towers for Super Power TV," *NAB Engineering Conference Proceedings*, National Association of Broadcasters, Washington, DC, 1988.

Wright, R.H., and Hyde, J.V., "The Hill-Tower Antenna System," *RCA Engineer*, August-September, 1955.

6.9 TELEVISION FIELD STRENGTH MEASUREMENT (54 MHz-806 MHz)

JOSEPH W. STIELPER, ANN GALLAGHER
MOFFET, LARSON & JOHNSON, INC. (MLJ) ARLINGTON, VA

INTRODUCTION

This chapter will discuss planning a program of television broadcast field strength measurements and the analysis of the resulting data. It is not intended merely as a cookbook describing the actual taking of television broadcast field strength measurements; most field strength meter instruction books are adequate for this purpose.

It is anticipated that many measurements will be made on digital television (DTV) signals that are presently beginning and will ultimately replace analog transmissions. DTV signal quality measurements will require new equipment and techniques; however, measurements of DTV field strength are currently made with standard analog equipment. The only additional changes are switching the meter's detector from "peak" to "average" and making appropriate adjustments to the data analysis to allow for the use of a meter designed to receive analog signals.

The advent of DTV has raised the question of the appropriate "metric" for television coverage. In this chapter, we will continue to use field strength, since the Federal Communications Commission (FCC) continues to define coverage in these terms. By definition, field strength is the strength of the electric field at the point of interest, typically, the receiving antenna, and is customarily measured in mV/m or μV/m. These values can be converted to equivalent plane wave power density (W/m^2 or mW/cm^2) by relating voltage to the impedance of the medium (120π or 377Ω). Similarly, by taking into account the effective area of the receiving antenna and transmission line loss, receiver input power can be calculated.

While the discussions in this chapter focus on television field strength measurements, similar techniques and methods apply to measurements in the FM broadcast band as well.

MEASUREMENT OBJECTIVES

Why take measurements? Television field strengths may be measured to accomplish several objectives. These may include:

- Determination of actual (as opposed to predicted) coverage contours or coverage of a particular area
- Evaluation of the performance of transmitting systems
- Measurement of spurious emission
- Evaluation of the effect of factors such as terrain and vegetation on coverage.

Other special studies that may require field strength measurements include those involving interference reduction or making frequency allocation decisions. The ongoing DTV measurement programs have sweeping goals generally different from those described above and hence are not described in detail here.

Measurements Versus Predictions

It is not necessary or even customary to take measurements to estimate the overall coverage of a station or the station's coverage of a particular area such as its city of license. There are numerous propagation models available that may be used to predict field strength and coverage. These models vary in complexity from the administratively convenient but overly simplistic FCC $F_{(50,50)}$ curves to complex models which trace rays along the path from the transmitter to one or more receiving locations. Currently, the *Longley-Rice* model is used for DTV allocation and interference studies. More recently developed models based on the geometrical theory of diffraction (GTD) may be better suited to this purpose.

The FCC Rules specify the propagation model that must be used to determine coverage for Commission purposes, although additional showings using alternate models may be made. The FCC Rules also contain measurement and analysis procedures that must be followed if the results of measurements are to be accepted by the Commission. For in-house use, there are no such constraints, and any propagation model or measurement scheme may be used. The engineer is thus faced with the choice between measurements or predictions and calculations. The choice is not always clear. Predictions may be made in the comfort of one's office, with no concern over equipment availability or weather. Measurements, carefully taken and interpreted, can determine the signal levels actually available. The choice hinges on the answers to the questions:

- Which is more accurate or reliable?
- Will the results be filed with the FCC?
- What are the relative costs?

To a greater or lesser degree, propagation models consider the terrain and the environment between the transmitting and receiving antennas in a simplified fashion. Complex terrain and environmental situations involving multiple obstacles and reflections are reduced to much simpler models with one or two diffracting obstacles, usually described as spheres, wedges or knife edges. Many prediction routines do not consider the effect of environmental clutter such as buildings and trees. Such clutter may be more significant than terrain, particularly at UHF frequencies. In many cases it is not clear which of the many propagation models is appropriate. Because predictions often yield contradictory answers, there is a tendency to conclude that taking measurements will resolve the problem. Indeed, the FCC generally prefers measurements to propagation predictions, especially in contested situations. However, unless care is taken in the planning and execution of the measurement program, the results may not be an improvement on predictions and may be misleading. In summary, the methodology of any prediction or measurement program must be critically reviewed before basing decisions on the results.

Table 6.9-2

Median Field Strengths Required for Various Grades of Service in the Absence of Interfering Signals

6.9-2(a) FCC Technical Standards Analog National Television Systems Committee (NTSC)						
	Ch. 2–6		Ch. 7–13		Ch. 14–69[1]	
Grade of Service	μV/m	dBμ	μV/m	dBμ	μV/m	dBμ
Principal City	5000	74	7000	77	10000	80
Grade A	2500	68	3500	71	5000	74
Grade B	225	47	650	56	1600	64
6.9-2(b). Analog NTSC Based on TASO Data.						
Primary	250	48	1400	63	7500	75
Secondary	50	34	200	46	630	56
Fringe	20	26	55	35	180	45
6.9-2(c). FCC Technical Standards for Digital TV.[2]						
Principal Community	25	28	63	36	112	41

[1] For UHF, coverage contour adjusted by the dipole factor.
[2] Digital TV coverage is characterized by a single grade due to its threshold effect and assumed single standard for reception.

Standards and Definitions

The quality of service is related to field strength by considerations of receiver sensitivity and noise figure, receiving antenna gain and transmission line loss and tolerable signal-to-noise ratios (S/N). The DTV planning factors used by the FCC in developing DTV standards are shown in Appendix A. The required fields for established grades of service vary with the frequency assignment. Table 6.9-1 lists the frequency bands employed by television broadcast stations. The bandwidth of each channel is 6 MHz.

Table 6.9-2(a) lists values of median field strength required for various grades of analog television service in the absence of interfering signals as established by the FCC's Technical Standards.[1] Interfering signals from other transmitters on the same or adjacent channels may limit service to higher values of field strength. Table 6.9-2(b) shows revised estimates of the fields required in the various television bands to provide acceptable grades of service based on the practical experience of operating stations and the findings of the Television Allocations Study Organization (TASO).[2] The TASO study has never been officially adopted by the Commission and may now be obsolete. The present FCC standards for analog television may also be obsolete, but the concepts have taken on practical meanings and are still useful.

A number of changes in the definition of television coverage grades based upon the TASO studies and other data have been proposed from time to time since the adoption of the present definitions. A review of the rationale for the analog VHF contour definitions and some potential revisions are presented in *FCC Report FCC/OCE RS 77.01.3.* Table 6.9-2(c) lists DTV coverage standards as adopted in the *Sixth Report and Order* in *MM Docket No 87-268* based upon the planning factors of Appendix A.

Table 6.9-1

Frequencies Employed for Television Broadcasting Service

Channel Number	Frequency (MHz)
2–4	54–72
5–6	76–88
7–13	174–216
14–69	470–806

Grades of service for analog television, as defined in Table 6.9-2, refer to the median field strength with respect to both location and time, at a receiving antenna of nine m (approximately 30 ft) above ground. The median value of field strength is the value exceeded for at least 50% of the time at the best 50% of the receiving locations. The presence of trees, buildings, and terrain irregularities often result in considerable variation in field strength from one location to another, even within relatively small areas.[4–9] The variation in field strength with location must be taken into account in measuring field strength as well as in specifying service.

At this writing, the FCC DTV rules define service in terms of F(50,90) (that is, the field strength at 50% of the locations, 90% of the time) and require the calculation of distances to a prescribed service contour. For UHF the coverage contour of the above table is adjusted by the *dipole factor.* The resulting contour in dBu is given by $41 - 20 \text{ Log } (f/615)$, where f is the center frequency of the TV channel in question. In the absence of measurements to determine time fading factors during the measurement period, the engineer will probably assume that the measurement represents the median (50%) value with respect to time. Thus, coverage measurements will require adjustment by an appropriate fading factor to reflect 90% of the time. For DTV allocation purposes, the *Longley-Rice* propagation model is used to predict field strength. In general, both coverage and interference are predicted. Although coverage is for 90% of the time, interference is predicted for 10% of the time as for analog interference. Interference is usually not measured and in the case of cochannel interference could not be measured simultaneously with standard field strength meters. For comparison of measured data with *Longley-Rice* predictions, comparison with the predicted median may be used unless measured fading factors are available.

The following are scoring grades commonly used for visual and aural signals. The visual scale follows the TASO six-point system.[2] The aural scale is taken from International Telecommunications Union (ITU-R) Recommendation 262. These first two scales are used for the subjective rating of one aural or visual signal. Comparative scales may also be used where appropriate. Comparative tests tend to yield data that is more reliable and repeatable than tests on one signal.

Table 6.9-3

TASO Picture Quality Scale

Grade	Name	Description
1	Excellent	The picture is of extremely high quality, as good as you could desire.
2	Fine	The picture is of high quality, providing enjoyable viewing. Interference is perceptible.
3	Passable	The picture is of acceptable quality. Interference is not objectionable.
4	Marginal	The picture is poor in quality, and you wish you could improve it; interference is somewhat objectionable.
5	Inferior	The picture is very poor but you could watch it. Definitely objectionable interference is present.
6	Unusable	The picture is so bad you could not watch it.

Table 6.9-4

International Telecommunications Union (ITU-R) Aural Scale

Grade	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Obviously, with minor editing, the TASO scale may be adapted for aural use. The definitions would aid in improving repeatability in evaluation of the quality of aural signals. The following is the seven grade ITU-R aural comparison scale which may also be used for visual signals.

Table 6.9-5

ITU-R Comparison Scale

3	Much better
2	Better
1	Slightly better
0	The same
−1	Slightly worse
−2	Worse
−3	Much worse

PLANNING A MEASUREMENT PROGRAM

Before the start of measurements, a plan should be designed to consider the objectives of the field strength survey. For example, the constraints imposed by a survey to determine coverage of a station are different and generally less restrictive than those imposed by propagation studies. Design of a program to evaluate the performance of transmitting systems is likely to be the most difficult and require even more careful planning than programs for other purposes. Suggestions for designing measurement programs for each of the above objectives are presented in the following sections as well as a discussion of the techniques of actual field strength measurement. There are several basic considerations affecting all types of measurement programs that should be addressed. These include the choice of antenna height for the measurements and allowance for factors such as weather that are beyond the control of the engineer.

Measurement Antenna Height

The FCC Rules require measurements using a 9 m (30 ft) receiving antenna height. The bulk of reliable data was taken at 9 m, and the broadcast coverage contours are based upon this height. Thus, a 9 m receiving antenna height is appropriate for data to be filed with the FCC, for direct coverage measurement or for direct comparison with most other data.

Making measurements at 9 m introduces a number of practical problems (including personnel safety) which are discussed in a later section. For in-house use, another antenna height may be appropriate. If, for example, it is desired to evaluate the signal strength available to typical suburban viewers, it may be more appropriate to use a lower antenna to simulate indoor, set-top antennas. Measurements taken at a lower height may be, if necessary, adjusted to reflect the standard 9 m height. However, the variation in field with height may not follow simple laws. Because of the uncertainty of the magnitude of the adjustment, the adjusted results are inherently less precise than measurements taken at nine meters. Adjustment factors are discussed more fully in a subsequent section.

Measurement of Coverage

The coverage of a broadcasting station and the technical quality of the service provided are determined, in part, by the strength of the received signal as well as other factors. Currently available methods of predicting field strengths and service contours of television stations are only approximate, and even the best methods of calculating field strengths often fail to take into account variations due to important local conditions. In general, for making decisions affecting a station's operation, the most useful determination of station coverage is provided by properly made field strength measurements. This section describes measurement programs for this purpose.

A measurement program to determine coverage may be laid out using radials as in the method described in the FCC Rules for propagation studies (Section 73.686(b)). For example, eight radials spaced 45° apart typically starting with true north might be utilized. Additional radials, if necessary, may be added to obtain more refined coverage data. Under the FCC procedure measurement locations are evenly spaced at 2 mile (3 km) intervals. The point selection process defined in the FCC Rules is primarily intended to insure statistical randomness, that is, to minimize the effect of unintended biases that may affect the data. It also assures that the measured propagation path is a true radial. For in-house coverage studies, there may be no need for the measurement locations to be exactly on a true radial route at random 2 mile (approximately 3 km) increments as required by the FCC Rules. Because many TV transmitting sites are located near the center of cities it is often convenient to take measurements along more or less radial roads. Selection of measurement locations need not be random but may be influenced by population density or the desire to obtain data in particular areas.

Because actual picture or sound quality is only partially correlated with field strength, this may be only part of the answer sought when there are questions concerning coverage. For example, multipath causes ghosts in analog television pictures, raises the noise floor in DTV and causes distortion in analog TV aural (FM) signals, particularly to stereo signals. Such effects can be determined accurately by subjective observations. Care must be taken to ensure that the receiving test equipment and environment represent actual conditions. Observations made with high gain receiving antennas at 9 m (approximately 30 ft) may lead to erroneous conclusions if typical viewer reception relies on low gain indoor antennas essentially at ground level. For analog signals, it is often useful to make picture and sound recordings in addition to the field strength measurements. Recordings tend to significantly reduce problems such as variations among observers, repeatability and inaccurate reporting of the type of impairment. Recordings of analog signals can be used to measure signal parameters such as the amplitude and delay of ghost signals in the lab without transporting necessary equipment into the field, although recording equipment may introduce some distortion.

Analog recordings or DTV measurements are usually made in conjunction with field strength measurements using mobile runs or clusters of points. Ideally, picture quality should be obtained for a known field strength, preferably representing the median at the location. However, it may be difficult to select a point of a mobile run corresponding to the median value. Thus, the beginning or end of a run is often used for convenience; in this case, the field strength at the recording point should be noted.

For DTV there is no impairment of picture quality caused by noise or multipath as long as the symbol error rate (SER) does not exceed a threshold value. Thus, picture recordings are generally meaningless except in non-steady-state conditions. For example, there may be cases where man-made noise occurs in bursts which are short in duration compared to the time required to complete DTV measurements. Picture observations and recordings can provide information on instantaneous impairment. At any given location the most important value is the available margin over the required signal to noise of the system. The margin determines the reliability at that location. A number of DTV parameters may be measured at a selected location. These include SER, S/N, noise floor and fade margin, including the effect of multipath on the noise floor.

The results of field strength coverage surveys of analog signals are customarily presented as contour maps, showing lines of constant median field strength which represent the outer limits of various grades of service. A typical map of measured television station coverage is shown in Figure 6.9-1. Methods of preparing contour maps are described in detail in "Analysis of Measurements to Depict Coverage."

Analysis of Measurements to Depict Coverage

If a measurement height of other than 9 m (approximately 30 ft) was used, the received fields must be adjusted to the field expected at a receiving antenna height of 9 m above ground to obtain the location of standard contours. An antenna height of 3 m (approximately 10 ft) is often used when the results will not be filed with the FCC. It has been common practice to assume that field strengths (basic units of V/m) increase linearly with antenna height, as indicated by classical plane earth propagation theory. For this assumption the relationship between the field, E_{30}, which would be expected at 30 ft (approximately 9 m) and the field E_H, measured at a receiving antenna height H_r is $E_{30}/E_H = 30/H$. For example, the ratio of the field at 30 ft (approximately 9 m) to the field at 10 ft (approximately 3 m) is $30/10 = 3.0$, or 9.5 dB.

The application of the linear height-gain function discussed above is recommended only in relatively smooth terrain. In rolling or rough terrain the following height gain factors were recommended by TASO to convert from 3–9 m (approximately 10–50 ft) fields.

Table 6.9-6
Smooth Rolling Unobstructed Hilly Rough

Channel	Smooth Unobstructed Terrain	Rolling Hilly Terrain	Rough Terrain
2–6	9.5 dB	8 dB	7 dB
7–13	9.5 dB	7 dB	5 dB
14–83	9.5 dB	5 dB	2 dB

Median fields, as established in accordance with a radial or modified radial procedure described above, are plotted as a function of distance from the transmitter, and a smooth curve is fitted through the plotted points. Figure 6.9-2 is a typical graph showing the plotted field strengths as a function of distance from the transmitter, together with the smooth curve fitted to the plotted points. The points may be plotted with

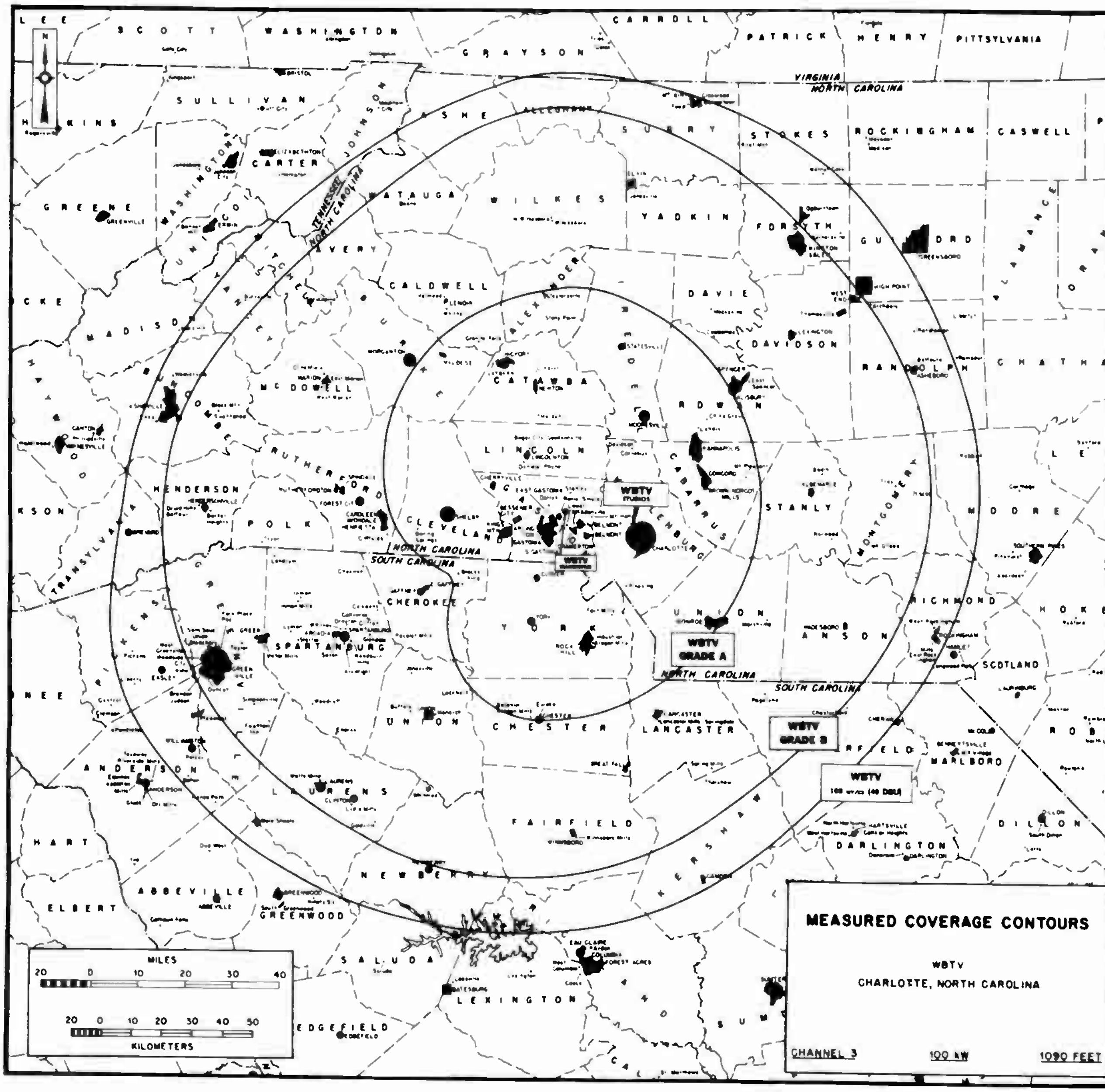


Figure 6.9-1. Map showing measured service contours for an analog operating television station. *(Courtesy of Jefferson-Pilot Broadcasting Company.)*

standard off-the-shelf spreadsheet software. Such software normally contains curve fitting routines. The solid curve is a simple linear fit to the measured data. Often, a first. second or third order polynominal is used to fit the data. The dashed curve in Figure 6.9-2 is the predicted field strength calculated using the propagation curves and prediction methods specified in the FCC Television Broadcast Technical Standards.[19]

Individual graphs of median field strength versus distance as shown in Figure 6.9-2 are prepared for each of the directions along which the measurements were made; the distances to the desired field strength contours, selected from Table 6.9-2 are then plotted on a suitable map and contours are drawn to produce a finished map such as shown in Figure 6.9-1. In some cases a mountain ridge may cause an abrupt decrease in field strength and define the limit of coverage. In this case, the curve-fitting procedure may not be appropriate. or only for distances less than the distance to the ridge. That is. data beyond the ridge should not be fitted in order to avoid incorrect bias of the data unaffected by the ridge.

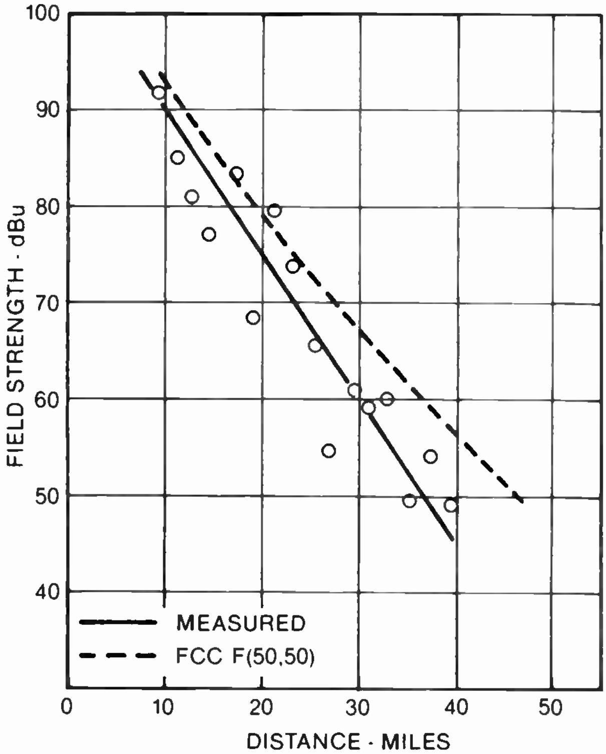


Figure 6.9-2. Graph of measured field strength versus distance for a typical radial series of measurements. The results of each mobile run are shown. The solid line is the *best fit curve* through the points. The dashed line is the predicted field strength from the FCC curves.

At present, DTV allocations are based on the *Longley-Rice* propagation model. In the case of existing stations, the goal is to replicate the station's Grade B contour on the assigned DTV channel. The DTV Effective Radiated Power (ERP) is derived at the station's antenna height above average terrain (HAAT) so that the F(50,90) DTV service contour replicates the F(50,50) analog Grade B. However, coverage and interference are based on predictions of field strength in cells (bins) 2 km (1.2 miles) on a side. There are approximately 8500 bins within the Grade B contour of a low VHF TV station with an antenna height of 1000 ft (300 m). Thus, it is generally not practical to measure field strength at even one location in each bin to verify predicted DTV coverage. Predicted DTV coverage using the *Longley-Rice* model is shown in Figure 6.9-3. Such maps also usually show predicted interference areas. Other than sampling at a limited number of locations, the only practical procedure would be to "drive test" with an antenna mounted on a vehicle roof. This procedure is described in the section on drive testing.

Transmitting System Evaluation

Field strength measurements are occasionally used to assess the performance of a station's transmitting system, particularly the operation of the antenna. The difficulty is in separating the effect of propagation factors from the effect of the transmission system. Ideally, comparison of transmission from the antenna to be tested should be made with that from a standard antenna. The standard antenna would be a low-gain antenna of known characteristics mounted near the antenna under test. This procedure is described in detail in the TASO report.[2] The procedure is rarely feasible.

- Measurements on two or more stations operating in the same frequency band are often used to determine if one is not performing properly. Differences in facilities and location must be considered in the analysis of the data. Vertical plane pattern effects may be minimized by avoiding locations near any of the transmitting sites or by adjusting data based on antenna vertical plane patterns. The uncertainty in the results is proportional to the number and degree of the adjustments that are made; numerous measurements should be taken to reduce the uncertainty.
- Comparisons of measured data with standard propagation curves such as the FCC Rules curves or smooth earth curves must be used with caution. The standard deviation of measured data used to draw the FCC curves is 7.7 dB at low VHF, 6.8 dB at high VHF, and 9.3 dB at UHF after adjustment for terrain.[10] Even in smooth terrain, field strength is affected by other factors such as trees, buildings, and atmospheric variations.
- In hilly terrain, the effect of clutter, terrain, and atmospheric variations can be minimized by selecting sites near the transmitting antenna to minimize atmospheric effects and ensure that ray path clearance over all obstacles can be obtained. This clearance should be at least 0.6 Fresnel zone. The radius of a full Fresnel zone is given by:

$$H_0 = 2280 \sqrt{\frac{d_1 d_2}{f(d_1 + d_2)}}$$

where:

H_0 = Fresnel zone radius in ft
d_1 = Distance from transmitting antenna to obstacle in miles
d_2 = Distance from receiving antenna to obstacle in miles
f = Transmitting frequency in MHz

Application of this criterion is illustrated by Figure 6.9-4. Even if such ray clearance over all obstacles can be obtained, reflections, especially from terrain, may often modify field strength. Allowance for reflections can only be achieved by using mobile runs and by taking measurements at a large number of locations to "average out" reflection effects.

Field strength measurements intended to evaluate system performance are sometimes taken in an aircraft in an attempt to reduce terrain effects. However, there are several problems associated with aircraft measurements, including:

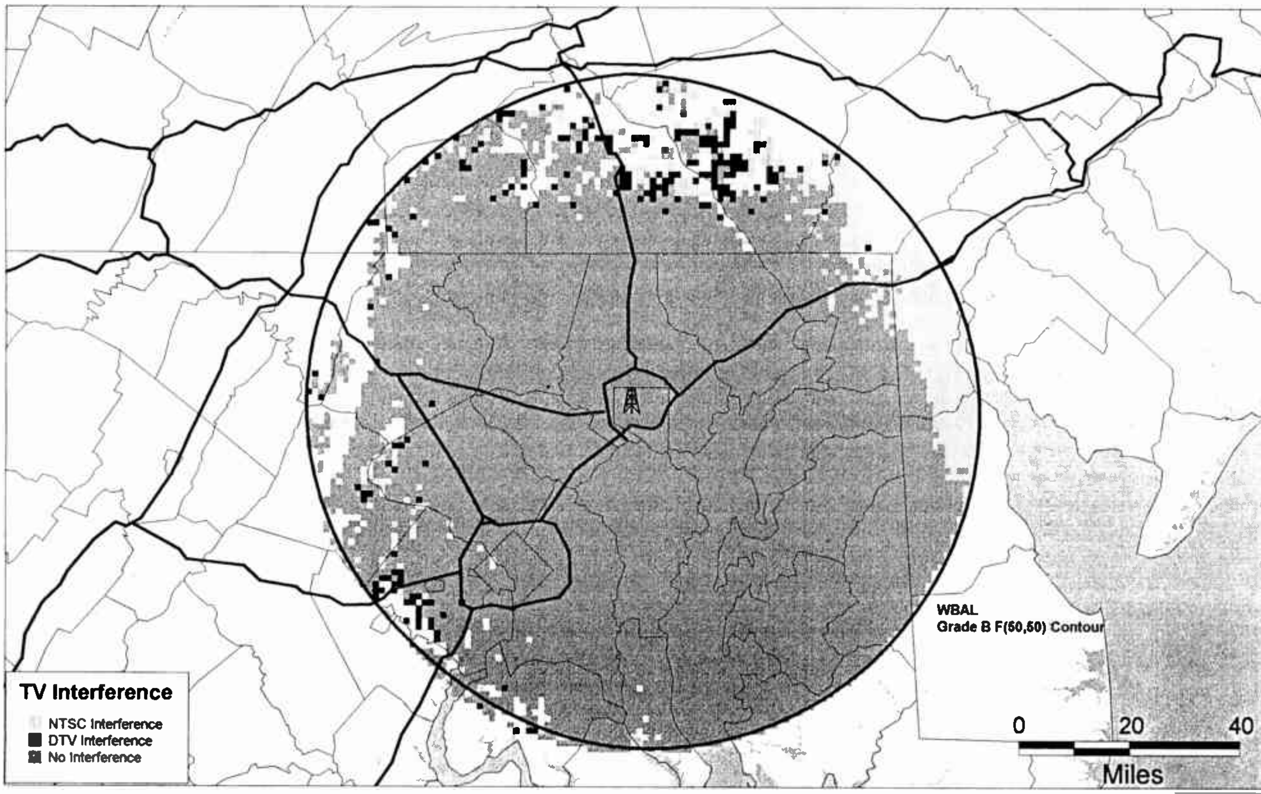


Figure 6.9-3. Predicted Coverage using the Longley—Rice propagation model and the replicated Grade B Contour.

- Determination of the actual location of the aircraft relative to the transmitting system under test
- Constraints on the design of the program imposed by the aircraft's performance
- Mounting of the receiving antenna on the aircraft to minimize the effect of the airframe on the receiving antenna pattern without creating a hazard

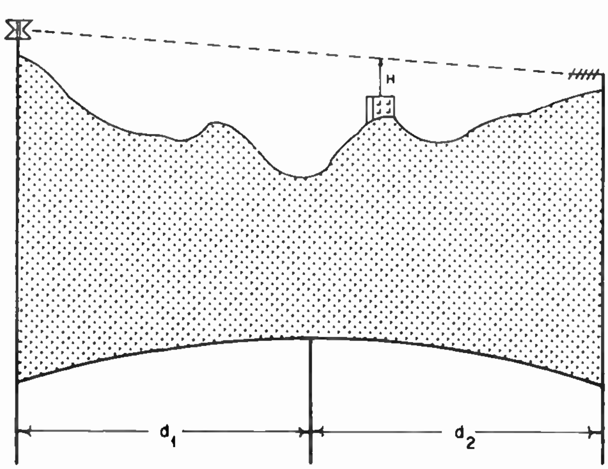


Figure 6.9-4. Illustration of required ray clearance to avoid obstacle loss.

- Uncertainties, such as actual path loss, that limit the accuracy of the measurement of absolute gain of the transmitting system under test.

The methods for determining the location of the aircraft or maintaining the desired location vary from simple manual techniques to control using electronic location measures such as Global Positioning System (GPS). In the most basic case, an altimeter is used to maintain the desired height. A more or less circular route can be flown by flying over known ground locations. At the other end of the scale a completely electronic system can be used. Such systems are used in Europe because proof measurements are routinely required after construction.

Reflections are a limiting factor in antenna measurement whether measurements are taken on a test range, on the ground or on an aircraft. In the latter case, the standard procedure is to fly close to the antenna under test so that the reflecting ground will likely be illuminated by a relatively low ERP. The distance should not be selected arbitrarily but should be based upon the vertical pattern of the antenna. Ideally, a helicopter should first be used to measure the antenna's vertical plane pattern at a number of locations to select the radius of the horizontal plane flight. If this is not feasible, a *range measured* or theoretical vertical plane pattern should be used. If possible, the radius should be selected so that the depression angle of a ground-reflected ray falls in a null of the vertical plane pattern. For the case of a well smoothed vertical pattern, the flight radius may not be optimized, but the engineer should estimate the uncertainty so that unwarranted validity is not attached to the data. Error estimates may be made based upon theoretical reflection coefficients in NBS Technical Note No. 101.1.[2]

If measurements are taken on circularly polarized signals, in most circumstances, care should be exercised to avoid flying at distances where the angle of incidence of the ground-reflected ray would be at or near the Brewster angle. At this angle, a vertically polarized wave changes phase and the magnitude of the reflection coefficient approaches zero. This would be desirable when measuring vertically polarized only signals to minimize the effect of ground reflections. Vertically polarized only transmissions rarely, if ever, occur in television. If the angle of incidence approximates the Brewster angle, it will be difficult to compare the results between the two planes of polarization of circularly polarized signals because of the differences in reflection coefficient between vertically and horizontally polarized signals.

Even if Brewster angle problems are avoided and an aircraft flight path is carefully controlled, ground reflections will likely cause apparent, but fictitious variations, in measured antenna patterns. For example, at an aircraft and antenna height of 300 m (984 ft) and a flight radius of 2 km (1.2 miles), a change of 10 m (approximately 33 ft) in ground elevation at a reflecting point or aircraft elevation changes the reflected ray path length by approximately 5 m (approximately 16 ft). This is more than one half wavelength at all TV broadcasting frequencies. Thus, relatively small changes in the elevation of the reflecting point or aircraft can completely change the phase of a reflected ray. In this example, if the magnitude of a reflected ray is only one-tenth of that of a direct ray, the expected variation caused by ground reflections is approximately ±0.9 dB. Because of the effect of ground reflections, measurements at varying distances should be taken to average out these effects.

Perhaps the most difficult aspect of aircraft measurements is the accurate determination of the gain of the receiving antenna as mounted on the aircraft. The aircraft and the antenna transmission line can change antenna gain significantly from the nominal "free space" value. Unless receiving antenna gain is accurately known, only relative antenna patterns can be measured. The best method would be to measure receiving antenna gain by measurements on transmitting installations of precisely known ERP.

Analysis of Measurements to Evaluate Antenna System Performance

The following description applies to data taken on land at locations where at least 0.6 Fresnel zone clearance is obtained over all obstacles; data from aircraft measurements may be analyzed in a similar fashion. For aircraft measurements, essentially continuous data is analyzed; in ground measurements, discrete data is used.

If a station uses an omnidirectional antenna, the data may first be analyzed by neglecting noncircularity. For a directional (or an omni) antenna with a measured or theoretical pattern and known orientation, measured data may be adjusted for the horizontal antenna pattern. Relative field strength may then be determined assuming free space propagation for comparison with the antenna's theoretical or range-measured vertical plane pattern. The result of such an analysis is shown in Figure 6.9-5. Data may also be analyzed by comparing measured to predicted field strengths for free space propagation. If the antenna is performing properly the median difference should be small and approach the accuracy of the test procedure. The median difference between measured and predicted strength for the data shown in Figure 6.9-5 is 1.2 dB that is, measured field strength exceeded predicted field strength.

A comparison of measured field strengths may also be used to analyze measurements on different signals. Adjustments to allow for fairly small differences in height and location can be made directly to the measured ratio of field strength. In decibels, a factor of 20(log(height ratio)) can be used. For example, to adjust for a height difference between 800–1,000 ft (approximately 240–300 m) a factor of 1.9 dB is added to the lower height data. For distance adjustment, a factor of 40(log(distance ratio)) may be used. In this case (for example for distances of 9 and 10 miles (or approximately 15 and 16 km), 1.8 dB would be added to the data at 10 miles (approximately 16 km). This procedure is best used on stations on the same frequency, such as comparison between a main and standby antenna. By using this technique it is not necessary to calculate predicted field strength. This procedure is less desirable than the use of the Fresnel zone method because of uncertainties in the propagation paths of the two signals and uncertainty in the adjustment factors. The adjustment factors are based on the plane earth propagation model which may not be apropriate at some locations, for example, at a location where free space propagation obtains.

PROPAGATION AND OTHER SPECIAL STUDIES

Section 73.686 of the FCC Rules describes the procedures to be employed for field strength measurements used in TV propagation studies. This procedure is basically the technique developed by TASO. The actual field strength measurement procedure is discussed later in this chapter. The FCC procedure is intended primarily to yield data that may be analyzed to study the effect of terrain and other local influences on field strength. Measurement programs intended for other purposes may require substantially different plans. For example, measurements primarily intended to determine the effects of clutter, such as trees and buildings, should be taken, if possible, in smooth terrain to eliminate the effect of terrain.

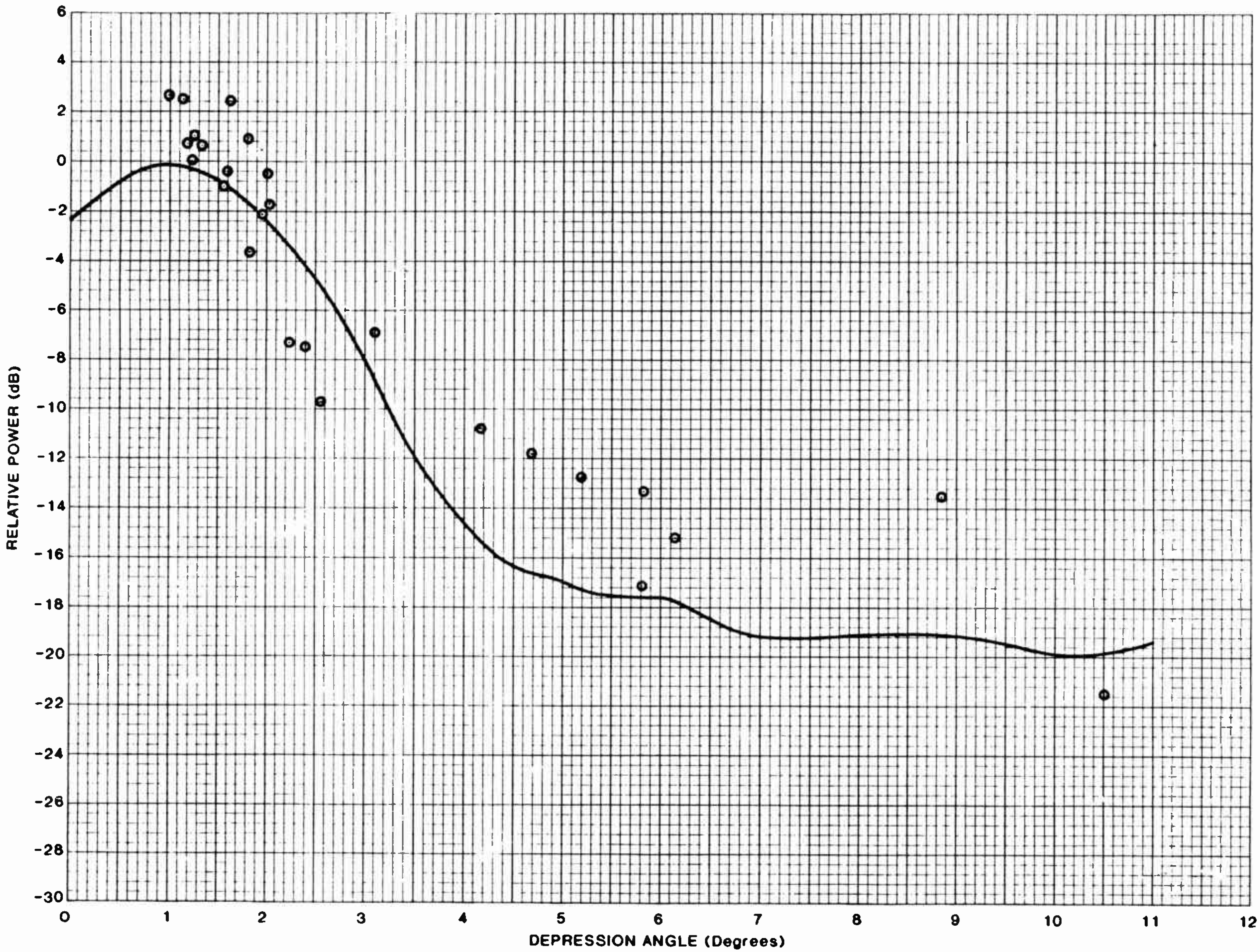


Figure 6.9-5. Comparison of data based upon field strength measurements with antenna vertical plane pattern.

Measured field strength for propagation studies may be compared to calculated field strength for a number of propagation models.[10–13] In this manner, the validity of the model may be assessed or factors for clutter may be determined. Measurement programs have been conducted to evaluate field strength time variation factors. In this case, field strength measurements are recorded at fixed locations over a period of time.

Field strength measurements may be an integral part of special field tests, particularly with regard to interference studies, allocation proposals and changes in TV operation. Examples of such field tests in the past include those on VHF TV-Land Mobile channel sharing, tests relating to the problem of educational FM interference to Channel 6 and on circular polarization for television. Extensive measurements were made in southern California in the study of propagation between Santa Barbara and San Diego regarding the proposed allotment of television Channel 10 to Santa Barbara. The presently ongoing DTV studies are also an example of a special program. Detailed discussion of such special projects is beyond the scope of this chapter; however, the basic techniques of field strength measurement discussed herein are valid for special programs. In addition, many of the topics discussed in planning coverage, transmission equipment evaluation and pure propagation tests are appropriate for use in special tests.

DTV Measurements

As discussed briefly above, DTV parameters can be measured; the special equipment needed is reviewed below. Data from the initial DTV field tests indicated good correlation between field strength and observations of picture quality for UHF transmissions although correlation was poorer at VHF. Subsequently, a protocol was developed for DTV field testing, and tests are proceeding using the protocol. A brief summary of the protocol follows. Measurements are made on arcs of constant distance from the DTV station and at selected locations on radial routes. In addition measurements are taken at "grid" locations which are determined by a grid laid out on a map. The radial and grid programs are similar to plans based upon the FCC Rules. All measurements are made with a 9 m (approximately 30 ft) antenna height and a directional receiving antenna directed towards the DTV station. Measurements of received signal level and DTV parameters are usually made at one point at each location although mobile field strength runs may be made. The segment error rate, equalizer output, signal and pilot levels and S/N are measured. To determine fade margin, white noise is added to the received signal. Multiple measurements are taken when a low margin is measured to determine if improvements can be achieved. These include measurements after changes in antenna height and antenna orientation as well as moving the test van.

Measurement of Spurious Emission

It is occasionally necessary to measure spurious radiated field strength from TV broadcast facilities to show compliance with FCC Rules after installation. Requirements regarding the level of spurious signals relative to carrier level are specified in Section 73.687 of the FCC Rules for TV transmitting systems. Such field strength measurements are nominally regulated by Section 2.993 of the FCC Rules. Because this section is quite vague, it is recommended that the person intending to take measurements determine acceptable procedures from FCC laboratory personnel before undertaking the measurements.

Drive Testing

In the mobile radio services, "drive testing" is often used to measure coverage and to conduct propagation studies, particularly to "optimize" propagation parameters for the prediction of coverage and interference. Drive testing has also been performed to a limited extent in FM broadcasting. In drive testing, a receiving antenna is mounted on a vehicle roof and continuous measurements may be taken on any road. There are a number of problems with TV drive testing. Most TV transmissions are, or, in the case of DTV, will be horizontally polarized. The receiving antenna must be omnidirectional and an omnidirectional, horizontally polarized rooftop antenna is difficult to design, particularly at the lower VHF frequencies. The vehicle will affect the antenna pattern particularly in the vertical plane. However, most DTV allotments, especially during the transition period, are in the UHF band. At UHF, design of a horizontally polarized omnidirectional antenna for vehicular operation should be feasible. The sensitivity of such a system will be lower than that of a measurement system that employs elevated directional antennas. In other words, for a given transmitting system, the range measured by a drive testing system will be shorter than that of a conventional measurement system. As in any other measurement program, care should be exercised in the design of a drive testing program so that the reliable data are gathered and to ensure that biases are minimized. As described in the section on practical field strength meters, it is necessary to insure that noise does not contribute significantly to the field strength indication. Two other potential sources of bias result from driving on streets approximately radial from the transmitting site and on major highways such as interstates. In these cases, the effect of clutter is generally much less than in residential areas where TV viewing occurs.

The antenna height on a vehicle in drive testing will, in many cases, be more representative of actual heights in a community and the data may be more useful for coverage measurements than will measurements at 9 m (approximately 30 ft). Drive testing may be helpful in evaluating potential power increases and comparative coverage from different sites where stations are presently operating.

Data obtained by drive testing may be used for comparison with predictions based on various propagation models and to derive *clutter factors.* Generally to derive *clutter factors,* an attempt should be made to characterize receiving environments; there may multiple environment types in a survey area. Data may then be segregated by environment, and factors may be determined for various environmental types. In any case, calculations of predicted field strength should be performed for the antenna height used in the survey to avoid height gain adjustments.

BASIC EQUIPMENT PRINCIPLES

Field strengths in the VHF and UHF bands (30–3000 MHz) are ordinarily measured by determining the voltage which the field induces in a half-wave dipole. The basic relationships can be expressed in several forms. The power transferred between two half-wave dipoles in free space separated by a distance, d, is given by:

$$P_r/P_t = 2(1.64)(\lambda/4\pi d)^2 \qquad [2]$$

where:

P_r = Received power
P_t = Transmitted power
λ = Wavelength in same units as d
d = distance between transmitting and receiving antennas

In terms of the field at the receiving dipole, the power delivered to a matched load by a half-wave dipole in a field of E volts/m is

$$P_r = (0.0186E\lambda)^2 \text{ watts} \qquad [3a]$$

where λ is expressed in meters.

Or, alternatively, in dB relative to one watt (dBW)

$$P_r = F - 20\log f - 105.1 \qquad [3b]$$

where f is the frequency in MHz and the field strength, F, is in dBu.

For a resistive load of R ohms, the voltage V developed across a matched load by a dipole in a field E is

$$V = (E\lambda\sqrt{R})/53.7 \text{ [volts]} \qquad [4]$$

The fundamental problem presented, therefore, is that of measuring the developed radio frequency (RF) voltage by a practical instrument of acceptable accuracy.

The voltage measuring device is ordinarily separated from the antenna by a length of cable. The cable introduces loss, and any impedance mismatch must be sufficiently small that calibration errors are not introduced by differences between the antenna and cable impedance and the internal impedance of the calibrating oscillator.

The following paragraphs describe the basic components used for field strength measurement using the standard FCC methods. The system described is the conventional system using chart recorders and manual analysis of recorded data. Digital sampling may be used to replace, entirely or at least in part, manual techniques. Discussions of the procedures for use of both chart recorders and digital recorders are included in order to aid engineers who may prefer either method.

Practical Field Strength Meters

Field strength meters are calibrated receivers that fall into three basic categories:

- Receivers that contain a precision oscillator and attenuator and use direct comparison between meter readings produced by the received power and the output of the oscillator
- Receivers that use a precision oscillator to adjust receiver gain to produce a direct reading meter
- Receivers that are direct reading but do not contain a calibrating oscillator.

Field strength meters of Type 3 are not considered suitable for precision field strength measurements. In addition, spectrum analyzers may be used for field strength measurements; the use of such instruments is discussed below.

Figure 6.9-6 is a block diagram of a practical Type 1 field strength meter. Type 2 meters are more common now; however, the purpose of the following discussion is to illustrate measurement principles. The Type 1 meter is more suitable for this purpose. The antenna delivers its received power to a transmission line leading to the receiver input. If the receiver input is unbalanced to ground, a balance-to-unbalance transformer (balun) is required. The transmission line between the antenna and the receiver is shielded to avoid stray pickup.

The RF attenuator shown serves two purposes: to avoid overloading of the receiver input on strong signals and to improve the impedance match when the receiver input impedance is substantially different from the characteristic impedance of the transmission line. It is frequently omitted when not required for either of these purposes.

The signal at the receiver input is amplified and converted to the intermediate frequency. Amplification and attenuation at the intermediate frequency permit operation over a wide range of field strengths; further range is provided by the receiver gain control. The rectified receiver output operates the indicating meter.

In operation, the attenuators and gain control are adjusted to provide an on-scale reading of the indicating meter. The receiver input is then switched between the output of the transmission line and the output of the calibrating oscillator, which is tuned to the frequency being measured. The output of the calibrating oscillator is adjusted to a predetermined fixed value using the RF power monitor, and the calibrated attenuator is adjusted until the indicating meter deflection is the same as that obtained from the antenna and transmission line.

For this condition, the voltage at the output of the calibrated attenuator is the same as that from the antenna and transmission line. By taking line and balun losses into account and applying 4 above, the field at the antenna required to produce this voltage can be determined. The relationship between field strength and receiver input voltage is usually expressed as $E = KV$, where K is a function of frequency. Figure 6.9-7 is a typical graph showing values of K for a VHF field strength meter.

A typical commercial field strength meter of professional quality is shown in Figure 6.9-8. The instrument

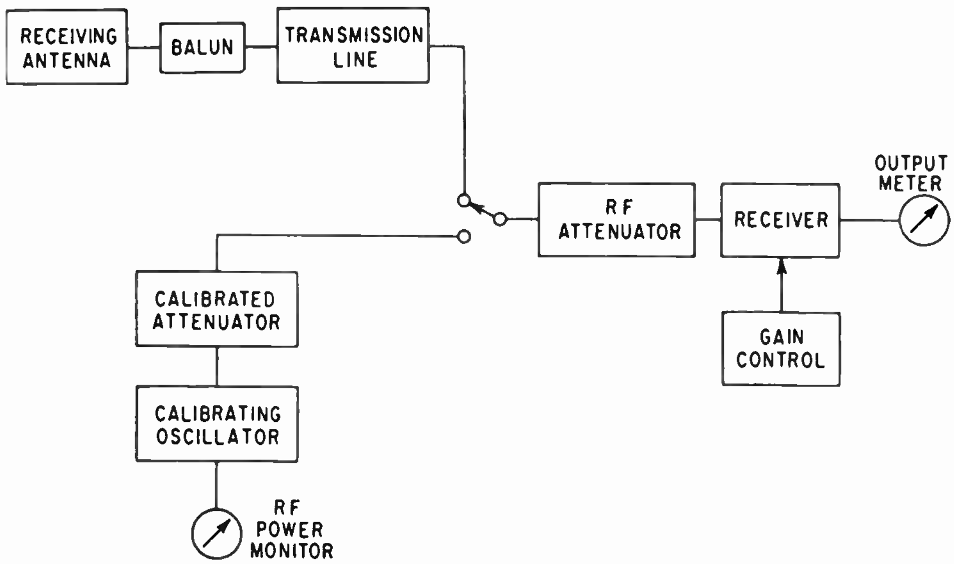


Figure 6.9-6. Block diagram of practical field strength meter.

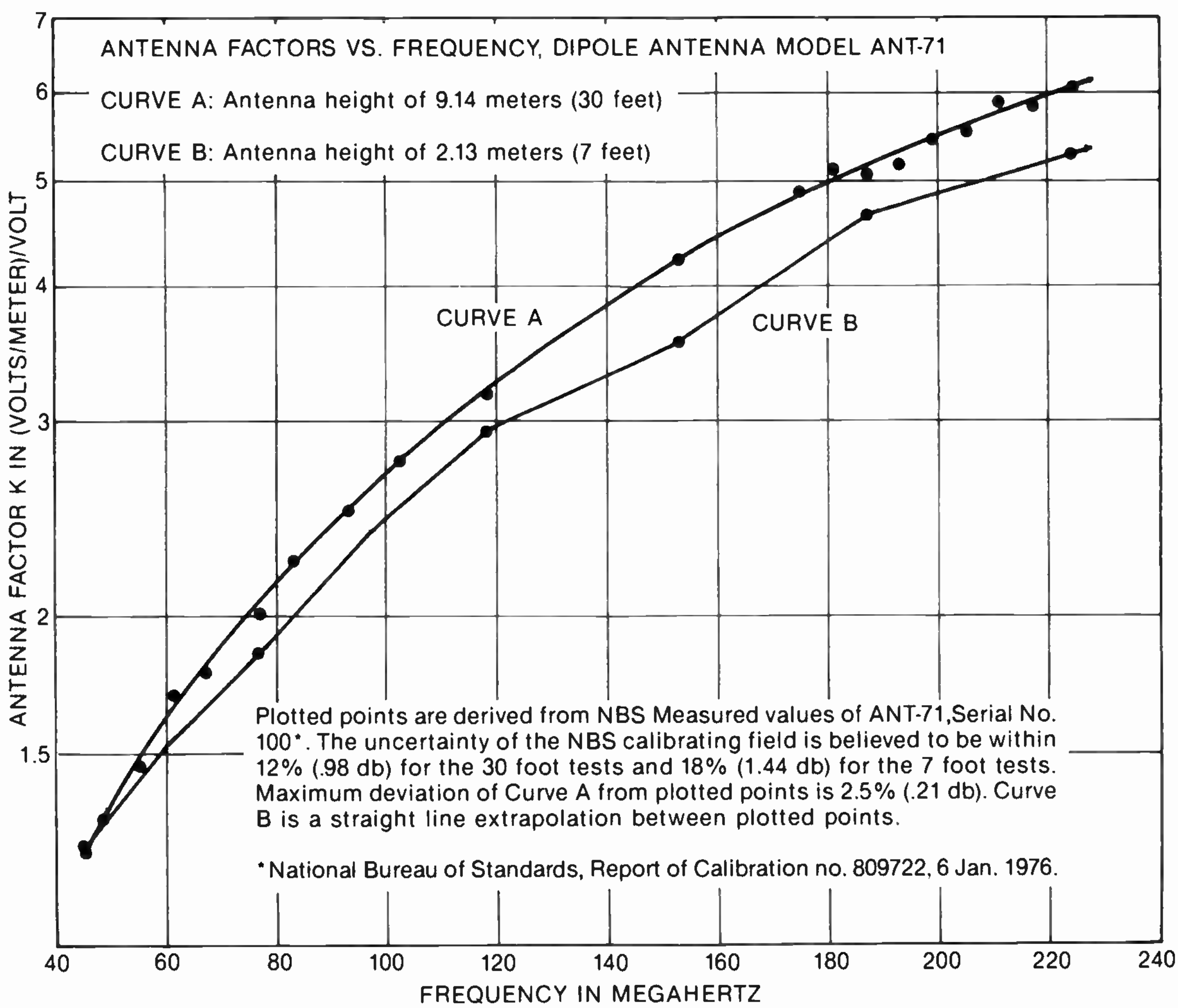


Figure 6.9-7. Graph of K for a typical VHF field strength meter. *(Courtesy of Potomac Instruments, Inc.)*

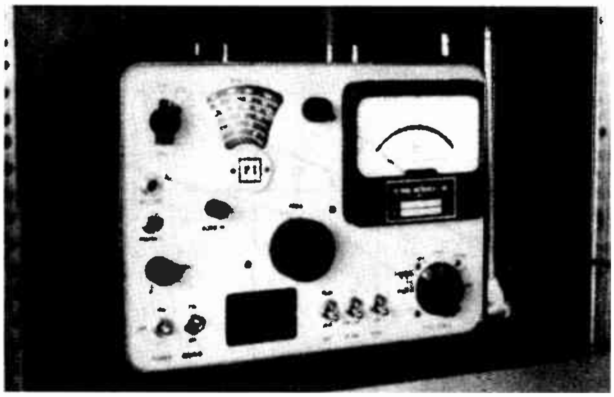


Figure 6.9-8. A VHF meter of professional quality.

shown is a Potomac Instruments type FIM-71, covering the VHF, FM, and television band from 54–216 MHz. A companion instrument, the FIM-72, which is similar in appearance, covers the UHF television band from 470–806 MHz.

Accurate instrument calibration is essential in measuring RF fields. During use, the calibration of the instrument described is provided by the calibrating RF voltage source, which is usually an integral part of the field strength meter (see Figure 6.9-8). The calibration of the oscillator and the calibration of the instrument as a whole must in turn be established and maintained by reference to laboratory standards.

The most direct laboratory calibration of the complete field strength meter is established by generating a known standard field in which the receiving antenna is placed. Standard field ranges have been developed and constructed at both UHF and VHF and are sometimes used in primary calibration of field strength meters.[14,15] Most commercial laboratory calibrations, however, are made by removing the dipole elements from the standard antenna and applying a known RF voltage at the proper frequency to the dipole terminals in series with an impedance equal to the receiving-antenna impedance. The calibration of the balun, line and receiver is established in terms of this applied voltage, which is then related to field strength through Equation 4.

Figure 6.9-9. Equipment set up for field strength measurement.

The calibration of the internal reference oscillator section includes the calibration of both the oscillator proper and the variable-output attenuator, if employed. The attenuator is usually of the inductively coupled piston type,[16] which depends only on its dimensions for proper functioning. This can be checked against the correct dimensions or against a laboratory standard attenuator. The oscillator can be compared with a standard oscillator, or its output can be measured with a laboratory standard such as a bolometer bridge.[16] This calibration is normally, but not necessarily, performed by the manufacturer.

Measured Signal Power

If measurements are made on the visual carrier of an analog television station, the difference between the peak and average powers of the transmission must be taken into account. This can be done by establishing a calibration in terms of average power for a still scene (such as test pattern or black picture), or a peak-reading meter can be employed to indicate the level of the synchronizing peaks. Such peak-reading meters are an integral part of professional commercial field strength meters such as the one illustrated in Figure 6.9-8. For DTV measurements, the standard is average power, so that peak reading is not required.

Spectrum analyzers may be used to replace standard field strength meters to permit the simultaneous measurement of multiple signals. If the multiple signals are not arriving from a common direction, for example, from an "antenna farm," an omnidirectional receiving antenna must be used. In the case of analog television measurements, care must be exercised to ensure that peak of sync is measured. At fixed points, the "peak hold" feature may be used to measure peak of sync. The total power of an FM signal is constant, so the level of an analog TV aural carrier may be used if a peak detector is not available. However, the power of an FM carrier diminishes under modulation. To obtain accurate measurements (particularly mobile measurements) on aural TV FM signals, the spectrum analyzer must be adjusted to "see" the bandwidth of each signal. In addition, intermodulation signals may be generated within the spectrum analyzer itself. If there is any doubt, the validity of signals should be checked.

Equipment for Field Strength Measurements

In addition to the field strength meter, several accessory items are needed when making a field strength survey. The principal items and their use are described in the following paragraphs and include:

- A special receiving antenna
- An antenna supporting mast
- A chart or digital recorder
- Power supplies.

Figure 6.9-9 shows the field strength meter, chart recorder, and some additional equipment for a survey.

The size and weight of the equipment usually dictate that it be mounted in an automobile or light truck. As discussed in Appendix A, all of this equipment, including the field strength meter, should be grounded to the vehicle frame. Figure 6.9-10 shows a van with elevated mast supporting a UHF antenna and containing the mounted equipment of Figure 6.9-9.

In addition to taking field strength measurements, it is often desired to use other equipment such as the television monitor shown in Figure 6.9-9. A vehicle devoted to general field test programs should have sufficient space available for special equipment including monitors, waveform scopes and magnetic recorders.

For DTV measurements, standard TV field strength meters are operated in the average power mode. In addition, an adjustment must be made to allow for the difference in bandwidth between the meter and the DTV signal as well as other characteristics of meters designed for analog measurements. Presently, bandwidths of the meter must be adequate for measuring peak of sync for analog signals, approximately 0.5 MHz, whereas for DTV received power is averaged over the full 6 MHz channel. The adjustment factor for the meter shown in Figure 6.9-8 is approximately 12 dB taking into all factors into account when the meter is tuned to the center of the TV channel. That is, true DTV received voltage is 12 dB greater than indicated. Engineers should consult the manufacturer of the meter to obtain the appropriate correction factor. Adjustment factors can be determined for individual meters by measurement; the manufacturer of the meter should also be consulted in this case. As can be seen from Table 6.9-2, DTV coverage standards are approximately 20 dB lower than corresponding analog values. Even allowing for differences in modulation, field strength meters designed to measure analog signals may lack the necessary sensitivity to measure DTV signals near the coverage limit. Use of directional receiving antennas can generally compensate for this problem. A low noise amplifier (LNA) can also be used to improve sensitivity. The LNA should be mast-mounted ahead of the transmission line if an elevated antenna is employed. The specifications of the meter coupled with receiving antenna gain and line losses, can be used to calculate system sensitivity. The system sensitivity should be several dB lower than the minimum signal to be measured, particularly at UHF. The noise floor should be determined by noting the meter reading when the antenna input is terminated with its characteristic impedance. The contributions of the meter noise can be eliminated by assuming power addition between noise and signal and adjusting the measurement downward. Alternatively, measurements less than a margin above the noise floor could be ignored; 6 dB is generally an acceptable margin. The error introduced would be less than 1 dB.

The equipment used to conduct specific DTV measurements can also be used to measure field strength. This equipment is described in a later section. Present DTV measurement equipment is not necessary nor particularly suitable for mobile field strength measurements.

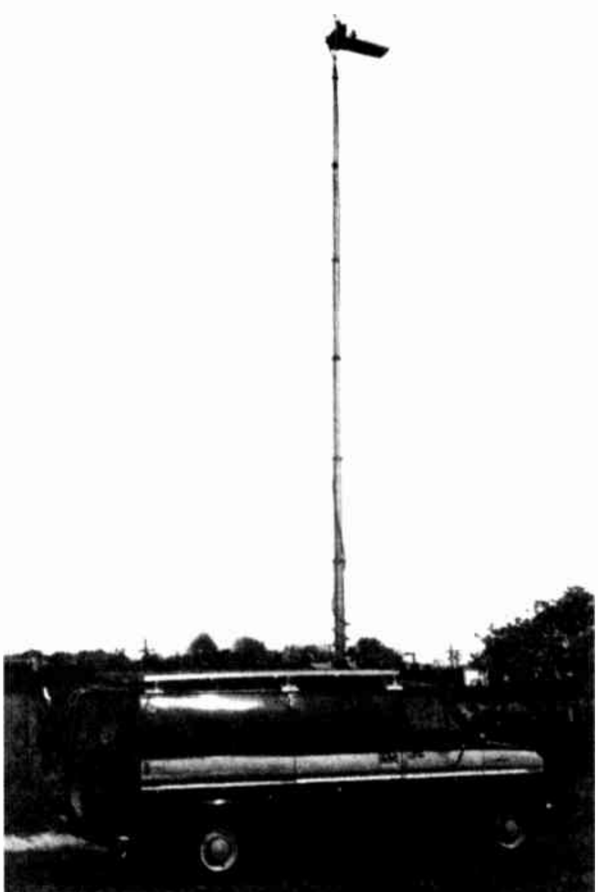


Figure 6.9-10. Van equipped for making field strength measurements with 30 foot antenna height.

Receiving Antennas

The measurement survey can be made by employing the standard dipole antenna furnished with a field strength meter, or other antennas can be utilized. It is often desirable to use an antenna other than the standard antenna for actual measurements. Standard antennas are usually not mechanically sturdy. If measurements are to be taken at a large number of locations, especially with extensive driving between locations, it may be more practical to use a rugged or readily replaceable antenna and thus minimize the probability of damaging the calibrated antenna. For subjective evaluations of analog signals or measurement of DTV parameters such as delay spread, the antenna should be representative of antennas actually used in the area by the public.

Where measurements are to be taken on several channels at each location, use of standard antennas requires that the length be adjusted for each channel within a band. It is also necessary to change antennas for VHF and UHF bands. Broadband antennas such as log periodics can significantly speed up the measurement process by eliminating the need to change antennas. Broadband antennas must be calibrated against the appropriate standard antenna to be useful.

An antenna that is essentially omnidirectional in the horizontal plane does not require orientation as the vehicle is moved; however, it is more common to use a receiving antenna with directivity. Directional receiving antennas possess gain which is especially useful for UHF measurements, and their directivity is useful to eliminate unwanted interfering signals. Directional receiving antennas also tend to discriminate against reflected signals, thus reducing apparent variation in measured field strength. A directional antenna can also be used to improve system sensitivity.

The antenna employed for the measurements must be calibrated on the measurement vehicle because of ground and vehicle proximity effects.[18] The differences in the calibration curves of Figure 6.9-7 are caused by ground proximity effects. The received field is first measured using the standard dipole antenna mounted on the vehicle at the measurement height. The antenna to be used in making the survey is then mounted on the vehicle, at the height to be employed in making the survey, and the receiver input voltage is determined with the receiving antenna at the same spot in the field. The procedure should be repeated several times and the median or average gain should be used. If an omnidirectional receiving antenna is used, the circularity of the pattern of the antenna as mounted on the vehicle must be determined. The antenna pattern is best established by rotating the vehicle with the antenna mounted as for measurements, and recording field strength and measuring antenna gain as above. If the vehicle cannot be rotated, an alternate comparable procedure must be followed.

The gain of a directional service antenna can be established relative to a dipole antenna by means of measurements with the antenna stationary as described above, but more consistent results are often obtained by making short mobile runs over identical paths and recording the signals from the two antennas. In either case, a location essentially free of standing waves should be used. For either procedure, the voltage gain of the service antenna G_s, relative to the standard dipole antenna G_d is $G_s/G_d = V_s/V_d$, where V_s and V_d are the voltages delivered to the receiver input using the service and standard dipole antennas, respectively. The power gain in dBd is 20 Log(G_d). If the transmission line or balun between the antenna and receiver is different from the standard cable and balun supplied with the instrument, the antenna calibration must include the measurement system cables and baluns even if standard equipment is used. Ideally, actual transmission line loss is measured and used in field strength determination.

Horizontally polarized receiving antennas are normally used for field strength measurements. At the present time, FCC Rules require that TV stations operate with horizontal only, or dual or circular polarization. If measurements on a vertically polarized field are desired, precautions must be taken to reduce the effect of coupling between the vertically polarized antenna and the supporting mast and vertical run of transmission line. In the case of horizontal polarization, the transmission line and mast are at right angles to the receiving antenna and thus exert negligible effect on reception. For vertical polarization, the line and mast are parallel and generally close to antenna elements. The receiving antenna pattern and gain may therefore differ significantly from free space pattern and gain. The standard antenna furnished with the field strength meter is used for measurement or calibration of another antenna, preferably directional. This measurement antenna may then be mounted close to the mast, and the gain is determined compared to the standard antenna which is offset from the mast. The standard antenna should be mounted at least one and preferably two wavelengths away from all metallic vertical components. This requirement presents a mechanical problem at the lower VHF frequencies. This measurement antenna is then calibrated against the standard antenna cantilevered out from the mast about two wavelengths. Thus, in the vicinity of the standard antenna, the transmission line is at right angles to the antenna. Since the mast and vertical run of transmission line are relatively distant from the antenna, the calculated variation in gain is only about ±0.5 dB when the standard antenna is cantilevered out by two wavelengths. The result is that the effect of the mast and transmission line on the measurement antenna is compensated for in the calibration.

Normally, directional receiving antennas are oriented for maximum field strength at the beginning of a mobile run. A GPS system and associated software can be used to determine the azimuth to the pertinent station to orient the receiving antenna. This is particularly useful when reflections cause ambiguous indications of the direction of the maximum. If GPS is not available, moving the vehicle can often change the effect of reflections and permit satisfactory orientation. In some cases, the direct signal can be maximized and orientation in the direction of the station achieved by orienting the antenna approximately in the direction of the station. A short mobile run can then be made to maximize signal strength. The antenna can then be "peaked" at location where the direct signal is strongest, and effects of reflections are minimal. In some cases, a reflected signal may be the maximum and orientation towards the station may not be appropriate.

Antenna Supporting Mast

The receiving antenna is ordinarily supported at a height of 2–9 m (approximately 6–30 ft) above ground, depending on the measuring technique employed. For the short heights, a simple mast of metal or plastic tubing can be used. For the standard 9 m (approximately 30 ft) height, a special pneumatic mast is normally used to raise and lower the antenna, and the mast arrangement should permit the vehicle to move over limited distances with the mast elevated.

The measuring unit shown in Figure 6.9-10 employs a telescoping mast typically constructed of aluminum tubing elevated by compressed air or nitrogen. The mast descends under gravity when the pressure is released. A handle inside the vehicle permits the mast to be rotated to orient the receiving antenna. Electronic news gathering (ENG) vans are normally equipped with a telescoping mast. Such masts may have tilt and pan heads that may be used for antenna orientation.

Operation with an elevated antenna involves a number of safety hazards posed by the measurement vehicle. Power lines are the principal overhead obstruction of concern. The avoidance of these hazards, equipment grounding and operation after encountering a hazard are discussed in Appendix A. These safety procedures were taken from the TASO Report[2] with minor modifications that update the procedures and description.

Recorder

For measurements made with the vehicle in motion, a chart or digital recorder is employed. A chart recorder can be driven from the vehicle speedometer or a clock drive motor. Excitation of the recorder is provided by a dc amplifier, which usually is built into the field strength meter or may be a separate accessory.

When the chart recorder is employed, the recorder pen element must be calibrated against the receiver output indicator of the field strength meter. The dc recorder is adjusted for balance at the ends of the meter scale, and a calibration curve is prepared for intermediate values.

Instead of a chart recorder, an analog-to-digital (A/D) converter can be used to sample field strength to permit computerized analysis of the measured data. A/D converters are available off-the-shelf enable the user to set the sample rate and generate recorded values. Portable laptop computers can be used to store and process the measurement data from the converter. Software can be included to analyze the data from a run to report the statistics of the run. The rate should be set sufficiently high so that the median, minimum, maximum and variance of the run data can be determined with precision.

Ideally, sample rate is a function of the frequency of the signal being measured. For reflection from a single source, the shortest wavelength of a standing wave is one-half of the actual wavelength. To provide reconstruction of the wave pattern to determine maxima and minima, a minimum of two samples per standing wave should be taken. For example, if channel 69 (the worst case) is measured the wavelength is approximately 1.2 ft (approximately 37 cm) and a sample should be taken every 0.3 (9 cm) ft or less (2.4 km per hour or 67 cm per second). Thus, if a vehicle travels at 1.5 miles per hour or 2.2 ft per second at least 7.3 samples per second should be taken. This is well within the capability of available A/D converters. However, the response time of field strength meters, including the recorder output, may be relatively slow which is not a problem when using chart recorders. Thus, increasing sample rate cannot compensate for excessive vehicle speed. To make matters worse, multiple reflections will tend to decrease the wavelength of standing waves. The worst case occurs when the field strength does not contain a direct constant component and consists entirely of reflected components. For an infinite number of reflections the resultant field is Rayleigh distributed. TV signals are not normally Rayleigh distributed, particularly for low VHF when an elevated antenna is used. There is a strong constant component of field strength so that field strength follows a Ricean distribution. Ideally, more study needs to be done to specify recommended sampling rates.

The issue of length of the mobile run arises naturally from the foregoing discussion of sample rate. The 100 ft (30.5 m) length of the rules is an arbitrary value developed originally by TASO. In some cases, it may be too short to determine the local median, let alone maximum and minimum values. The rules specify the length as of *at least* 30.5 m (100 ft) (emphasis added). Thus, it is important when writing test protocols not to limit the run length to 100 ft (30.5 m). With a chart recorder or other graphic display, the engineer can observe the recordings. With an A/D converter a running average can be reported so that the engineer can determine if the average is stabilized and is not significantly increasing or decreasing. If the average has not stabilized run length should be increased, if feasible. In some cases on a 100 ft (30.5 m) there will not be even one cycle of the local standing wave and thus the run should be extended if feasible. Usually such cases are at low VHF where the standard deviation is low and there is inherently more certainty regarding the data. Often, more cycles of the standing wave pattern can be observed when running toward or away from the transmitting site. A run oriented directly toward or away from the transmitting site should be used when feasible. The length of a run to determine a local mean was determined for Rayleigh distributed signal (worst case) for mobile radio signals.[29] For an error of ± 1 dB, a run length of 40 wavelengths is required. This criterion is only met for 100 ft (30.5 m) runs for frequencies in the UHF band. Application of this criteria to Channel 2 implies that a run length of approximately 700 ft (approximately .2 km) should be used. TV signals do not normally have a Raleigh, but rather a Ricean distribution as discussed above. In addition, the standard deviation of measurements of low VHF signals is less than that of mobile signals. Although the 700 ft (approximately .2 km) value is an overestimation, more study is required to determine minimum length for mobile runs. In summary, runs should be at as low a speed as is practical, should be as long as feasible, considering safety of the personnel and the observed data and should employ a high sample rate.

Power Supplies

The power drain of the measuring equipment can be fairly substantial, especially if much accessory equipment is employed. It is usually preferable to provide a power source for the measuring equipment separate from the vehicle battery. This may consist of a separate battery bank to operate the meter and accessories, or a separate 115 V ac alternator may be mounted in the vehicle. The latter is employed to operate the ancillary equipment shown in Figure 6.9-9.

DTV Equipment

Equipment such as antennas and transmission lines used to conduct DTV measurements are essentially the same as those employed in analog measurements. If field strength only is desired, analog field strength meters adjusted as described above may be used. Presently there is no equipment available designed expressly for DTV measurements, including field strength and other parameters.

Presently DTV measurements employ a vector signal analyzer, such as an Hewlett Packard type HP 89441A, to conduct measurements. This instrument permits both time and frequency domain measurements, as well as amplitude and phase measurements. S/N or noise threshold measurements may also be made, generally with an external noise source. This analyzer is relatively insensitive, and low noise amplifiers are generally used at the input to improve sensitivity. This instrument is extremely expensive and thus its use is not expected to be widespread.

MEASURING PROCEDURES AND TECHNIQUES

The FCC TV Technical Standards prescribe methods for making measurements to be submitted to the Commission. These or similar methods are also usually used in making other surveys such as measurements of station coverage. Variations from the official procedure are frequently introduced, some of these variations are discussed under earlier headings. The following paragraphs summarize the present requirements of the Commission's Standards.

The individual measurements consist of short mobile runs (at least 30 m (approximately 100 ft) along the road) at each location so chosen, with the receiving antenna at the 9 m (approximately 30 ft) height. If measurements are made on multiple stations for comparison purposes, it is desirable to mark the beginning or end points for each run to ensure that the run is made over identical paths on each channel. Before making the measurement run, the gain of the field strength meter is adjusted according to the calibration procedure for the meter that is used. The recorder output of the meter may also be checked to verify that the meter and chart recorder or A/D converter are functioning properly.

The chart or other recorder is used to record the field strength meter output, and the median, minimum and maximum values of the field for each recording are determined from the recording. When using a chart recorder precise determination of the median is usually made after completion of the survey. It is also useful to use the meter's calibrating oscillator to make a trace of an estimated median value on the chart paper after each run. The median of the run is estimated and the output of a precision attenuator may be adjusted to the estimated median. This provides an additional calibration check and permits a rapid preliminary calculation of field strength that is often useful to monitor the progress of a survey. Thus, if measured values out of the expected range are obtained, the engineer may cease taking measurements and thoroughly check the measurement system.

Figure 6.9-11 shows a sample of a typical chart recording obtained by this method. Certain field strength meters do not include built-in adjustable attenuators such as shown in Figure 6.9-8; however, there is usually an oscillator output which may be fed through an external attenuator to the input to achieve this calibration. The median, maximum and minimum values from the chart are then converted to received field strength by combining the individual calibrations of the antenna, transmission line, field strength meter, dc amplifier and chart recorder as discussed above.

When using an A/D converter to yield digital records, software may be developed to permit such checks after each run so that measured minimum, maximum and median field strength are known during the survey.

FCC Standard Method For the Collection of Propagation Data

The following discussion summarizes the FCC procedure in making field strength measurements for propagation studies. The forms for recording and submission of data and other technical requirements are presented in Section 73.686(b) of the Rules.

The Commission's Technical Standards require field strength measurement surveys to be made with mobile equipment along at least eight radial lines from the transmitter. The radials need not be laid out along bearings separated by 45° (beginning with true North) as is standard for contour prediction. Measurements are required to be taken from 16 km (approximately 10 miles) in increments of 3 km (approximately 2

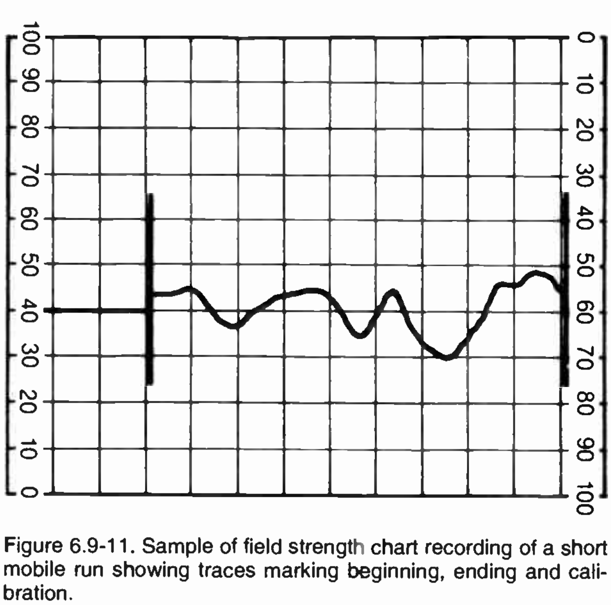


Figure 6.9-11. Sample of field strength chart recording of a short mobile run showing traces marking beginning, ending and calibration.

miles) in each direction. If it is desired to establish contour location, the distance should extend somewhat beyond the field strength contour which the engineer desires to establish. The routes are selected to encounter representative terrain and to permit interpolation between adjoining radials. A precise radial line is laid out from the transmitter on topographic maps to the distance to which measurements are to be made. Along this radial line, measuring locations are marked at exact 3 kilometer (approximately 2 mile) intervals, beginning at exactly 16 km (approximately 10 miles) from the transmitting antenna. The actual measurements are made precisely on the radial, at locations as close as possible to the exact 3 kilometer (approximately 2 mile) marks established as described. The ground elevation of the actual measurement location should be approximately the same as that of the intended location.

FCC STANDARD METHOD FOR MEASUREMENT OF COVERAGE OF COMMUNITIES

Section 73.686(c) of the FCC Rules describes the procedures for the measurement of service to specific communities for TV stations. These rules outline a measurement pattern that is in the form of a rectangular geographic grid overlying a map of the community. Measurements are made at locations as close as possible to the intersecting points on the grid.

The number of measurement points must be at least 15 or $0.1\sqrt{P}$ (whichever is greater) where P is the population of the community. Additional requirements describing documentation and calibration are contained in the rules.

The Rules also contain a statistical procedure to analyze field strength. This method fits the measurement data to a normal distribution and yields the median or average field strength in a community. The analysis may be graphical or numerical. This result is not completely compatible with other FCC Rules. The principal city coverage rules, either analog or digital, for example, are based upon the determination of the location of a coverage contour. This difficulty can be eliminated by using a grid that is not truly rectangular but consists of radials and orthogonal arcs. This plan permits both radial and grid analysis although there may be too few points on a radial for a good radial analysis. Since the resulting grid is not perfectly rectangular, use of this plan should be cleared with appropriate FCC personnel before undertaking measurements when the results will be filed with the Commission.

Practical Problems Encountered in Making Field Strength Surveys

Before any field strength measurement survey is undertaken, the radiated power of the transmitting installation must be established as closely as possible. The transmitter output power should be determined by means of the dummy load and maintained as closely as possible to the proper value throughout the survey. The radiated power is established from the measured transmitter output power, taking into account the antenna power gain and the transmission line and diplexer losses.

The use of a 9 m (approximately 30 ft) receiving antenna mounted on a vehicle in motion on a public right-of-way requires special permission from police or highway authorities in most states. These requirements vary among the individual states, but full details can be obtained from the state police or highway headquarters in the various state capitals.

The operation of such a 9 m (approximately 30 ft) mast presents safety hazards which require the exercise of utmost caution. In addition to proper grounding (discussed earlier) the TASO field strength measuring specification includes a special appendix (included here as Appendix B) dealing with overall safety requirements. When measurements are made with an elevated antenna, the need for caution must be borne in mind at all times.

Time Variation of Field Strength

Fairly substantial variations in field strength with time are frequently noted, particularly near and beyond the radio horizon, although significant variations occur at other distances.[10-12] These variations may be relatively rapid, occurring over a period of a few minutes, or slow variations may appear over periods of several hours. Field strength variations also follow diurnal patterns. There are also long term seasonal variations. In some areas during certain seasons there are cycles repeating over intervals of several days that apparently follow weather patterns. Average field strengths in the United States are usually lowest during winter afternoons, and higher than average fields may be observed during the evening hours and during summer. The variations in field strength with the passage of time must be taken into account in planning and making field strength coverage surveys.

The observed fluctuation of the field near the horizon is believed to be due principally to variations in the refractivity gradient of the lower atmosphere, which in turn is determined by the temperature, humidity and barometric pressure gradients. Measurements for coverage surveys should not be made during changing weather conditions or if weather fronts are known to be in the area.

The variations in field strength with time often result from causes which are not readily apparent and it is frequently difficult to determine whether typical propagation conditions prevail. One method that has been proposed and tried with some success is that of establishing fixed recording stations in one or more directions, at locations near the expected outer limit of the measurement program and recording the received signal over a period of several days or longer, if feasible. These recordings will give an indication of the signal to be expected under average conditions; the coverage survey measurements particularly near and beyond the horizon can be made during a period when recordings indicate propagation conditions to be typical. Measurements should not be made on days when these recordings indicate excessively high or excessively low field strengths. Measured data may be adjusted to reflect "average" conditions, although the adjustment factor would be a function of distance.

APPENDIX A DTV PLANNING FACTORS

System Independent Planning Factors Recommended by the Advisory Committee.

Planning Factor	Low VHF	High VHF	UHF
Geometric mean frequency (MHz)	69	194	615
Dipole Factor, dBm-dBu, $[K_d]$ (dB)	− 111.8	− 120.8	− 130.8
Thermal noise, $[N_t]$ (dBm)	− 106.2	− 106.2	− 106.2
Antenna Gain, $[G]$ (dB)	4	6	10
Download line Loss, $[L]$ (dB)	1*	2*	4*
Antenna minimum front to back ratio	10**	12**	14**
Receiver noise figure, $[N_r]$ (dB)	10	10	7
Time Probability factor, $[dT]$ (dB)	***	***	***
Location probability for 50% availability, $[dL]$ (dB)	0	0	0

* for 50 ft (15 m) of coax
** Ratio of forward gain to maximum response over rear 180° receiving antennas manufacturer's objectives are 14, 16, 20
*** The time probability factor is defined by the difference $F_{(50,10)}$ minus $F_{(50,50)}$; these two values are determined from the FCC charts in Section 73.699 of the Rules. This factor is a function of frequency band, distance and antenna height.

For the above quantities in [ ] and a required carrier to noise ratio of C_n, the required field strength (F) is then:

$$F = C_n - K_d + N_t - G + L + N_r + dT$$

APPENDIX B SAFETY

The following recommended safety precautions are based, in part, upon those developed by TASO for "Mobile Field Strength Measurements With Antennas Elevated 30 Feet Above Ground." When TASO formed, the field strength measurement procedure contained in the FCC Rules specified use of an antenna height of 10 ft (approximately 3 m) despite the fact that the 30 ft (approximately 9 m) standard for predicted fields was in the Rules. As discussed above, the field strength height gain factor for adjusting data from 10–30 ft (approximately 3–9 m) is not constant. Accordingly, TASO decided to use a 30 ft (approximately 9 m) height and a series of surveys were begun. It became apparent after a series of unfortunate accidents that the measuring of field strength at television frequencies can also be extremely dangerous where observations, especially mobile, are made at the 30 ft (approximately 9 m) level. The danger here is not with the field strength measuring equipment but with contact with primary electrical power circuits and, to a lesser extent, potential traffic hazards. In view of the dangers associated with field strength measuring, there are certain precautions that must be taken in addition to the specific precautions included within this memo for safety in operation of the field strength measuring vehicle. These general precautions include:

- Never take chances
- Always have two qualified, trained people present during measurement surveys
- Do not service or work on equipment when the power is turned on
- Never work on electrical equipment containing high voltages unless there are at least two persons present
- To reduce the risk of lethal currents across the heart, keep one hand in a pocket at all times
- Everyone who has occasion to work in the field on measurement programs should have a knowledge of a current certification in the best methods of artificial respiration and CPR.

Measurement Vehicle Construction Safety Considerations

Since it is possible that the mast or antenna of the measuring vehicle may come in contact with extremely high voltages, the construction of the vehicle should be such as to reduce to the absolute minimum the possibility of lethal differences of potential or voltages within the vehicle. The secure electrical bonds referred to herein must be made with a view towards the hundreds or thousands of amperes that may be involved in the event of an accident.

- All antenna elements which are directly connected to the transmission line must have a secure electrical bond to the mast, and the mast must be bonded to the vehicle chassis
- The outer conductor of all transmission lines used with the mast must have a secure electrical bond to the top of the mast
- The vehicle shall be equipped with appropriate warning lights to be operated when the mast is elevated
- The outer conductor of all transmission lines must have a secure electrical bond to the vehicle as soon after entering the vehicle as possible
- All electrical equipment (field strength meters, recorders, receivers, and signal generators) must have secure electrical bonding to the vehicle chassis
- The vehicle shall be equipped with a light readily visible to the vehicle driver which indicates that the mast is under pressure or elevated
- The vehicle should have a window in the roof to allow the driver or engineer, to view the mast and antenna.

The vehicle shall also be supplied with certain safety equipment, as follows:

- A pair of high voltage rubber gloves with protecting leather gauntlets. The rubber gloves should be tested at least once a year and a memo including the date of test and the testing organization included in the carrying box for the gloves. All major power companies have provision for making these tests
- A nonmetallic safety pole 8 ft (approximately 2 m) minimum for handling hot wires
- An appropriate fire extinguisher, preferably Halon, for Class A, B and C fires. This should be checked in accordance with the manufacturer's instructions.

Operational Safety Precautions

The foregoing equipment safety precautions are intended to enhance operator safety in the event of an accident. These precautions will only be effective as long as the equipment is in good working condition. Therefore, it is important that all safety precautions outlined in the foregoing paragraph be checked each morning before beginning the day's work. The following procedures apply when the mast will be elevated:

- Measurements at night with the mast elevated should not be permitted. In addition, the mast should not be raised during periods of poor visibility; small diameter power lines are often difficult to see. Although the authors believe that work at night and under other poor visibility conditions should not be permitted, TASO permitted night work under certain circumstances. The precautions were: No night work is permitted without prior written approval. If permission is given, it will include additional detailed precautions for the specific job for which approval is given. Prior to approving night work, the specific measurement locations must be surveyed during daytime hours for potential hazards
- The mast must not be elevated unless two operators are present
- The location for elevating the mast must be carefully chosen, both to prevent contact with overhead obstructions and to avoid being a traffic hazard. The mast must not be elevated on busy urban streets or heavily traveled rural highways. The chosen area must be reasonably level and, if a mobile run is contemplated, the vehicle must traverse the path before the mast is raised
- After selecting the location for measuring, the following procedure shall be used for elevating the mast. The driver and the engineer will both step out of the vehicle and examine the overhead area for obstructions. The engineer may then return to the vehicle and elevate the mast. If a mobile run is to be made, the driver must walk ahead, examining the path for overhead obstructions and leave a marker at the end of the chosen path. If repeated measurements are to be made along the same path, the starting point should also be marked so that neither end of the examined path is passed. The vehicle may then be driven between the markers
- When the measurements are completed, the driver shall, without stepping out of the vehicle, determine that the overhead area is free for the lowering of the mast. The driver must also make a personal observation that the mast has been fully retracted before driving to the next measuring area.

Emergency Procedure

If the procedures outlined here for the safe operation of field strength measuring are followed, there should be no need for emergency procedures. However, the authors felt it worthwhile to outline some emergency procedures to be followed in the event of some unforeseen accident involving overhead obstructions. Observe the following precautions:

- The obstruction with which you are entangled may carry high voltage
- The vehicle you are in may be at a high potential with respect to ground, so stop and think
- Under no circumstances should the transmission line (associated with the field strength measuring gear) be touched. The grounding connections may have broken due to mechanical strain or have burned up due to extremely high currents
- Without leaving the vehicle, study your predicament carefully to determine your best course of action. It may be one of the following:

  a. Backing the vehicle up
  b. Driving the vehicle ahead
  c. Lowering the mast
  d. Raising the mast further
  e. Getting away from the vehicle

Whatever you do, remember the vehicle may be at a high potential with respect to ground. Do not take chances. Jump clear. Be sure no one else approaches the scene or comes in contact with the vehicle.

ACKNOWLEDGMENT

The assistance of present and former associates Howard T. Head, J. M. Bixby and Eric-Paul Bomgren in preparation of this article, particularly this edition, is sincerely appreciated.

REFERENCES

1. Federal Communications Commission: TV Technical Standards, Secs. 73.683(a) and 73.685(a).
2. "Engineering Aspects of Television Allocations," Report of the Television Allocations Study Organization to the Federal Communications Commission, March 16, 1959.
3. Kalagian, G. S.:"A Review of the Technical Planning Factors for the VHF Television Service FCC/ OCE RS 77-01, March 1, 1977.
4. LaGrone, Alfred H.: "Forecasting Television Service Fields," *Proc. IRE*, vol. 48, no. 6, pp. 1009–1018, June 1960.
5. Brown, G. H., J. Epstein and D. W. Peterson, "Comparative Propagation Measurements; Tele-

vision Transmitters at 67.25, 288, 510 and 910 Megacycles," RCA Rev., vol. 9, no. 2, pp. 177–202, June 1948.
6. Bullington, K., "Radio Propagation Fundamentals," *Bell System Tech. J.*, vol. 36, no. 3, pp. 593–626, May 1957.
7. Peterson, D. W., and J. Epstein, "A Method of Predicting the Coverage of a Television Station," *RCA Rev.*, vol. 17, no. 4, pp. 571–582, December 1956.
8. Head, H. T., "The Influence of Trees on Television Field Strengths at Ultra-High Frequencies." *Proc. IRE.*, vol. 48, no. 6, pp. 1016–1020, June 1960.
9. Kinase, Akira, "Influences of Terrain Irregularities and Environmental Surroundings on the Propagation of Broadcasting Waves in the UHF and VHF Bands," Japan Broadcasting Corporation (NHK) Tech. Monograph No. 14, March 1969.
10. Damelin, J., W. A. Daniel, H. Fine and G. V. Waldo, "Development of VHF and UHF Propagation Curves for TV and FM Broadcasting," FCC Report No. R-6602, September 7, 1966.
11. Norton, K. A., "The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE*, vol. 29, December 1941.
12. Rice, P. L., A. G. Longley, K. A. Norton and A. P. Barsis, "Transmission Loss Predictions for Tropospheric Communication Circuits," NBS Technical Note No. 101, January 1, 1963.
13. Reed, H. R. and C. M. Russell: "Ultra-High Frequency Propagation," John Wiley and Sons, 1953.
14. Greene, Frank M., "Calibration of Commercial Radio Field Strength Meters at the National Bureau of Standards," National Bureau of Standards Circ. 517, December 1951.
15. Greene, Frank M., and Max Solow, "Development of Very-high Frequency Field Intensity Standards," National Bureau of Standards Research Paper RP2100, vol. 44, May 1950.
16. Terman, F. E., *Radio Engineers' Handbook*, McGraw-Hill Book Company, Inc., New York, 1943.
17. Schrack, R. A., "Radio-Frequency Power Measurements," National Bureau of Standards Circ. 536, March 16, 1953.
18. Greene, Frank M., "Influence of the Ground on the Calibration and Use of VHF Field Intensity Meters," National Bureau of Standards Research Paper RP2062, vol. 44, February 950.
19. Federal Communications Commission's Rules, Sec. 73.684.
20. Longley A. G. and P. L. Rice, "Prediction of Tropospheric Radio Transmission Loss—A Computer Method," ERL 79-ITS 67, Environmental Sciences Services Administration, July 1968.
21. Longley A. G., "Radio Propagation in Urban Areas," OT Report 78-144, U.S. Dept. of Commerce, April 1978.
22. Hufford, A. G. Longley and W. A. Kissick, "A Guide to the Use of the ITS Irregular Terrain Model in the Area Prediction Mode." U.S. Department of Commerce, April 1982.
23. Okumura, Y., Oiimori, E. Kawano, T and Fukuda, K., "Field Strength and Its Variability in VHF and UHF Land Mobile Radio Services," Rev. of Elect. Comm. Lab., vol. 16, no. 9–10, September-October 1968.
24. Keller, J. B., "Geometrical Theory of Diffraction," *Journal of the Optical Society of America*, vol. 52, No. 2, pp 116–130, Feb. 1962.
25. Luebbers, R. J., "Finite Conductivity Uniform GTD Versus Knife Edge Diffraction in Prediction of Propagation Path Loss," *IEEE Transactions on Antennas and Propagation*. vol. AP-32, pp. 70–76, January 1984.
26. Luebbers, R. J., "Propagation Prediction for Hilly Terrain using GTD Wedge Diffraction," *IEEE Transactions on Antennas and Propagation*, vol. AP-32, pp. 951–955, Sept. 1984.
27. Luebbers, R. J. et al, "Comparison of GTD Propagation Model Wide-band Path Loss Simulation with Measurements," *IEEE Transactions on Antennas and Propagation*, vol. AP-37, pp. 499–505, April 1989.
28. Chamberlin, K. A., et al., "An Evaluation of *Longley-Rice* and GTD Propagation Models," *IEEE Transactions on Antennas and Propagation*, vol. AP-30, November 1982.
29. Lee, William C. Y., "Estimate of Local Average Power of a Mobile Radio Signal," *IEEE Transactions on Vehicular Technology*, vol. VT-34, no. 1, pp 22–27, February 1985.

AUTHOR'S NOTE

Much of the present knowledge of wave propagation in these frequency bands has been derived from field strength coverage surveys on operational television stations. The information gained from these commercial coverage surveys has added to the body of scientific knowledge, but field strength measurement surveys employing special techniques are often needed to supply data for special problems. Examples of such special techniques are discussed under other headings in this article. It is better to design a program to measure specific factors, foliage loss for example, than to use general coverage measurements in an attempt to achieve the desired goal.

6.10
FIBER OPTIC TRANSMISSION SYSTEMS FOR BROADCASTING[1]

C. ROBERT (BOB) PAULSON
OMNIMEDIA® COMMUNICATION, WESTBOROUGH, MA

LIGHTWAVE SYSTEM DESIGN AND PRINCIPLE OF OPERATION

Signal transmission through either wired or wireless transmission media is accomplished by conditioning the input information signal to make it compatible with the characteristics of the transmission medium. The simplest system (Figure 6.10-1) transmits analog or digital electrical signals at *baseband* (the original signal waveform itself), typically through a copper based one or two-wire circuit. The input electrical signal amplifier on the receive end may contain an *equalizer* which changes received signal amplitude and/or phase as a function of frequency to compensate for the attenuation characteristics of the copper transmission medium (twisted pair of wires or coaxial cable).

A *modulator* is required to transmit signals electromagnetically through space, or through a glass or plastic fiber. Changes in the input signal amplitude are used to vary the amplitude, frequency or phase of a radio frequency carrier, or used to vary a light source's output intensity around a mid-level brightness (see Figure 6.10-1). Propagation of this light source into free space, or through light conductors, such as fiber optic material, follows essentially the same laws which describe the propagation of radio frequency electromagnetic radiation from radio wave antennas.

Signal transmission using fiber optic technology is accomplished by varying the flow of photons in a transparent optical fiber (typically smaller in diameter than a human hair), rather than the amplitude and frequency of radio waves in free space or the electrons in a copper wire. A basic fiber optic transmission system (Figure 6.10-2) consists of an optical transmitter terminal, a fiber circuit which can range in length from a short patch cord to up to 50 km of tandem fiber sections and an optical receiver terminal.

Design efficacy of a fiber optic transmission system is characterized by its ability to maximize the amount of light coupled into, propagated through, and coupled out of the fiber.

Semiconductors designed as optical transmission system light sources emit light energy in the infrared/visible light regions of the electromagnetic spectrum (see Figure 6.10- 3). At the transmitter terminal light source output, a tiny lens focuses its diverging light rays into glass fiber conductor (also called an *optical waveguide*). The signal carrier is essentially a monochromatic (or *single wavelength*) light source.

Inputs to the optical transmitter can be anything from analog sound (bandlimited to 25 kHz) to moving images with bandwidth up to 3 GHz and beyond. When digitized, the input signals' binary digital representations range from 56 kbps to 3 Gbps and beyond. The voltage modulating the light source may consist of one analog or digital electrical input, or many individual signals multiplexed together. The optical transmitter driver contains electronic processing circuit modules to condition (preemphasize) the input signal, and is followed by an *electrical* to *optical* (E/O) *transducer,* a complex, solid-state semiconductor device which converts electrical input energy into optical energy. Preemphasis is done to maximize system end-to-end performance. The unmodulated optical source output intensity (power) of the transmitter is expressed in dBm (decibels referred to 1 milliwatt of power) and is typically in the range from 0 to -30 dBm.

The group of optical rays delivered into the fiber constitutes the (intensity modulated) *carrier* of the information. The exact characteristics of these light rays propagated through a fiber are determined by the size and construction of the fiber. Maxwell's equations show that light does not travel randomly through a fiber; rather, it is channeled into *modes.* A mode is the path of one ray of a discrete wavelength through a fiber.

At the receive end, the attenuated intensity-varying light emitted from the fiber is focused onto a solid-state photo detector (photodiode) which transduces the light intensity variations back into an electrical signal. Typical threshold sensitivity of photodiodes (the input power level below which it can not differentiate the light signal from its own internal noise) is in the range of -25 to -35 dBm. Then, subsequent circuit components amplify, demodulate, reequalize and/or reconstruct the signal into its original form. These receiver terminal components are the complement of the transmitter's electrical input modules.

Conversion of an *electrical domain* input signal to an *optical domain* output signal is a complex process. Simply described, electron flow into the semiconductor causes a loss of input *bandgap energy,* which is released from the top surface of the semiconductor or at an edge as a steady state light at about 50% of maximum intensity.

[1] Some of the historical material in this chapter first appeared as Chapter 4.4 in the *NAB Handbook Seventh Edition*, authored by Pete Mountanos.

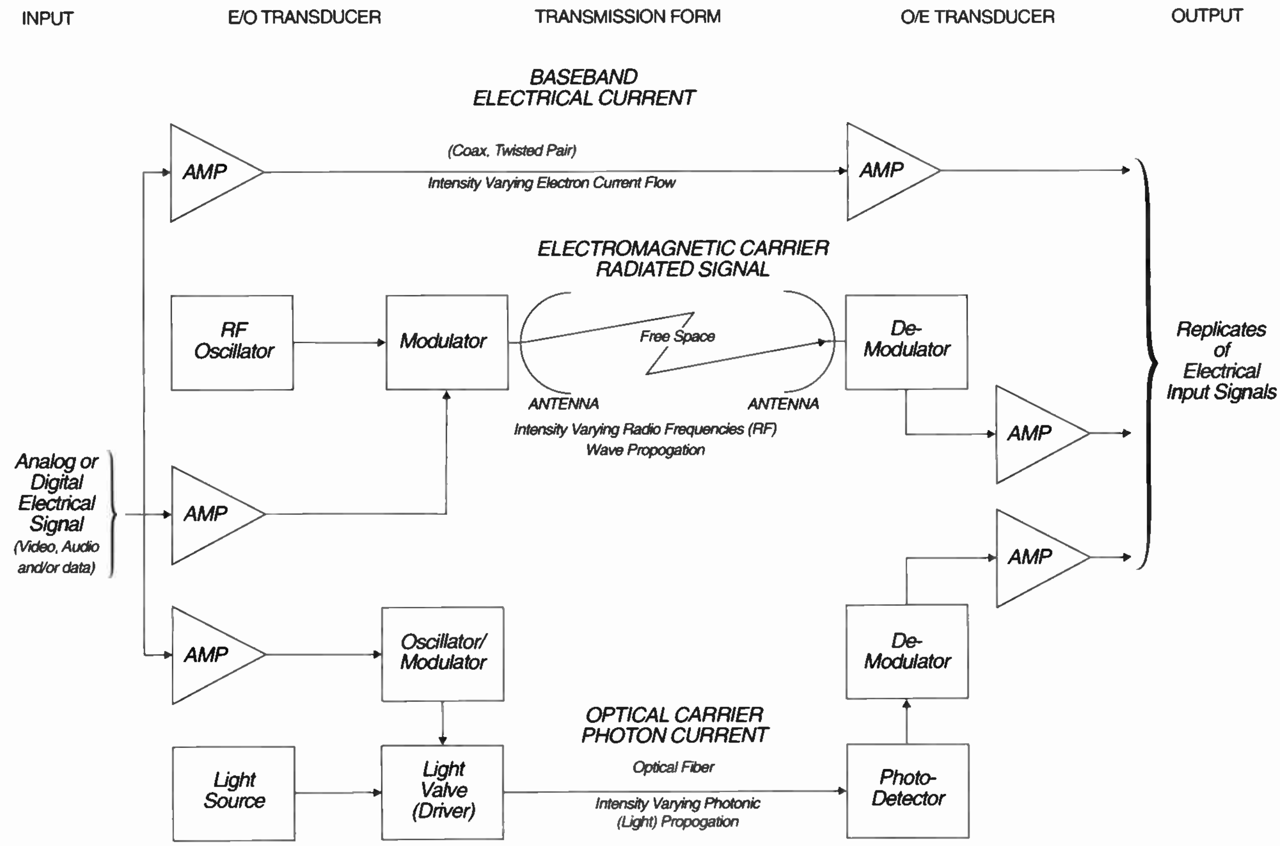


Figure 6.10-1. Means of electrical signal transmission.

Information carried as a time varying analog signal or a binary digital signal is processed into an electrical current flow, which time varies or switches the steady state electron current magnitude symmetrically above and below that value. This produces a corresponding intensity variation of the light output. However, the transducing process is not linear (see Figure 6.10-4). Therefore, the amplitude of the input electrical signal is processed by a limiter to prevent driving the output intensity into the nonlinear extremes of this transfer characteristic curve. This conservative approach limits intermodulation distortion (IMD) that would otherwise occur, but also establishes a maximum bound on signal-to-noise ratio (SNR) of the fiber system.

Light Propagation Fundamentals

Frequencies of light used in fiber optic transmission are in the range of 300–400 THz (10 × E14), several orders of magnitude higher on the electromagnetic frequency scale than the highest frequency radio waves (see Figure 6.10-3). Light waves are more commonly described in terms of wavelength (expressed in nano-

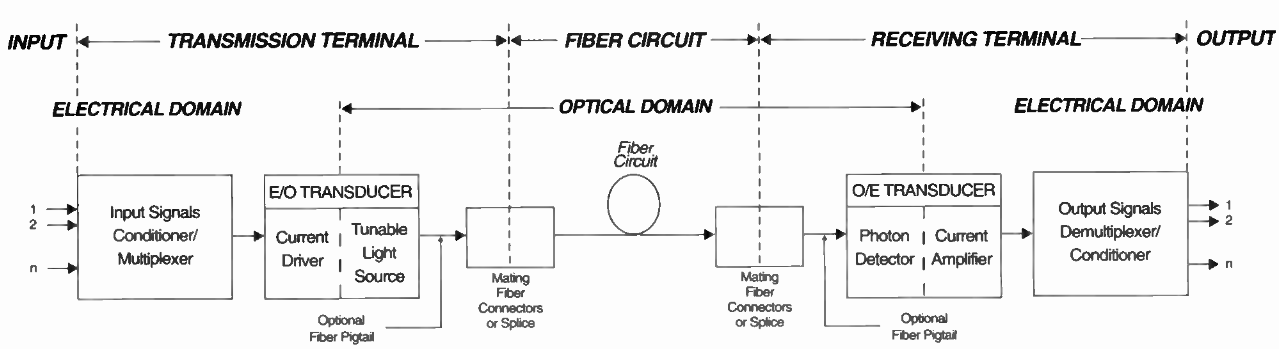


Figure 6.10-2. Terminal components in point-to-point (p-p) fiber transmission circuit.

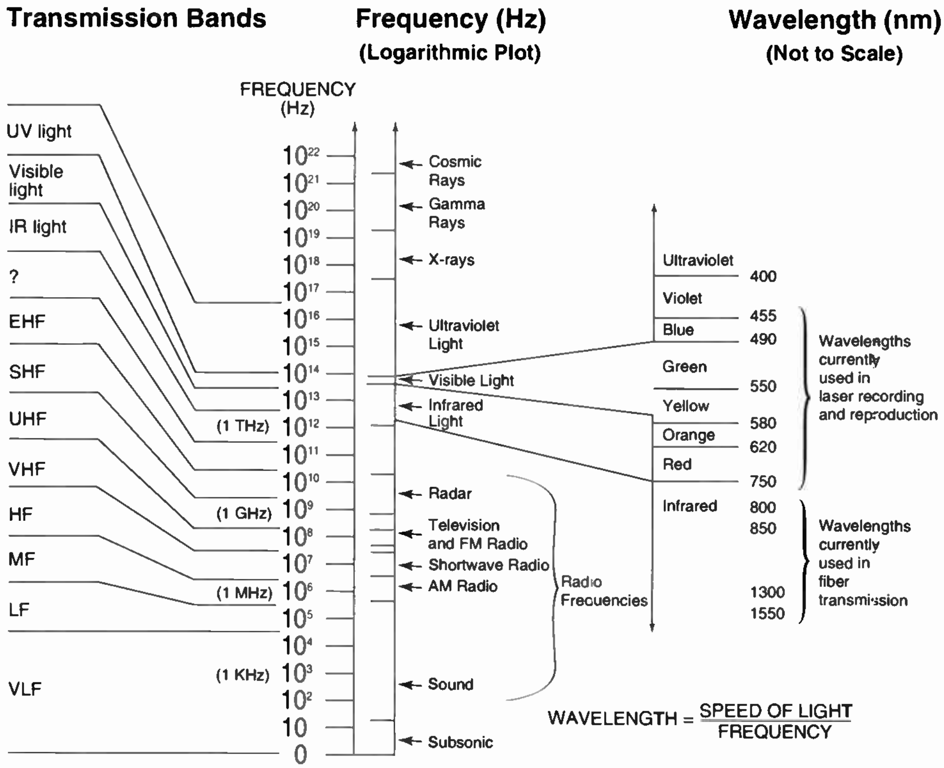


Figure 6.10-3. Electromagnetic spectrum (IR, visible and UV light spectra expanded). *(Adapted from AMP, Inc. Illustration with permission.)*

meters) rather than frequency. The region extending from 800–1600 nm is of greatest interest because today's glass fiber formulations propagate these wavelengths most efficiently (see Figure 6.10-5).

The speed (velocity) of light and all other higher and lower (electromagnetic) frequencies in free space is 300,000 km/sec. Light travels at measurably slower

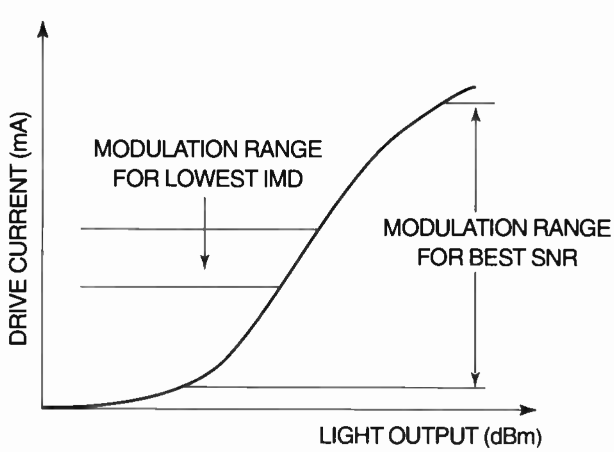


Figure 6.10-4. Drive current/light transfer characteristic of photodiode. A restricted modulation index is necessary when employing analog modulation techniques, to limit harmonic and intermodulation distortion products.

velocities in gas and solid media such as optical glass fibers. Each light wavelength propagates at different velocities in each medium (see Figure 6.10-6). When an electromagnetic wave crosses the boundary between one medium and any other, its propagation velocity changes. This is manifested as either a change in the direction (refraction) or reflection of the light wave.

The index of refraction "n" is a dimensionless number expressing the ratio of the velocity of light in free space (c) to its velocity in a specific medium (v)

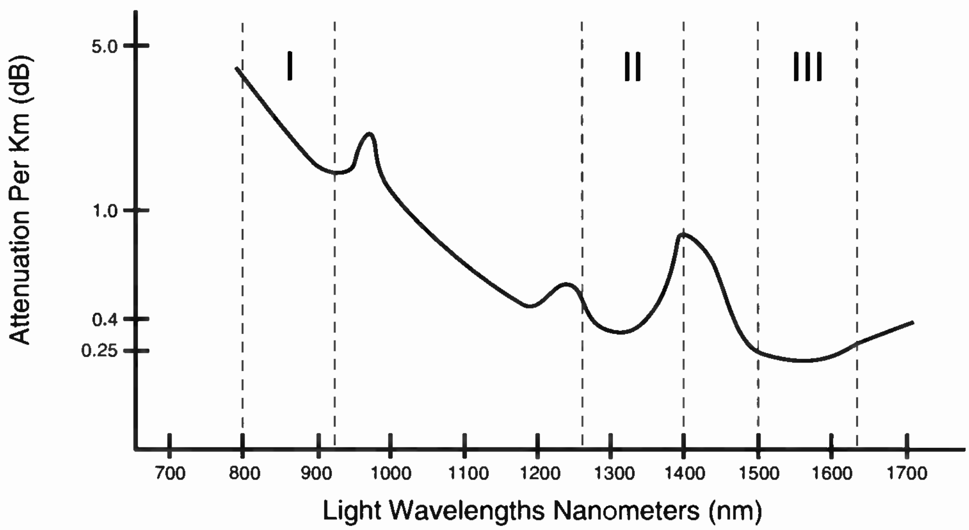
$$n = c/v$$

The magnitude of refraction of a ray of light as it passes from one material to another depends on the magnitudes (n_1 and n_2) of the refractive indices of the two materials (see Figure 6.10-7—reference normal is the line perpendicular to the interface of the materials). The angle of incidence is the included angle between the incoming ray and normal, and the angle of refraction the included angle between normal and the refracted ray. At small incidence angles, most of the light passes across the boundary and the transmitted ray is refracted away from the normal. As the angle of incidence increases to a *critical angle*, the angle of refraction is 90° from normal. For angles larger than the critical angle, the light is totally internally reflected (see Figure 6.10-7(c) and Figure 6.10-8).

A. Defacto Standard Light Wavelengths (Nanometers-nm)

Nominal Spectrum (nm)	Window	Fiber Types	
		MULTIMODE	SINGLEMODE
850 ± 30 *Short Wavelength*	I	X	
1300 ± 30 *Long Wavelength*	II	X	X
1550 ± 30 *Extra Long Wavelength*	III		X

B. Fiber attenuation versus light wavelength characteristics



C. Typical optical fiber attenuation ranges

Fiber		Optical Loss (dB/km)			
SIZE	TYPE	780 nm	850 nm	1300 nm	1550 nm
9/125 µm	**S M**	3.0	**2.5**	0.5 - 0.8	**0.2 - 0.4**
50/125 µm	**M M**	3.5 - 7.0	**2.5 - 6.0**	0.7 - 4.0	**0.6 - 3.5**
62.5/125 µm	**M M**	4.0 - 8.0	**3.0 - 7.0**	1.0 - 4.0	**1.0 - 4.0**
100/140 µm	**M M**	4.5 - 8.0	**3.5 - 7.0**	1.5 - 5.0	**1.5 - 5.0**
110/125 µm	**M M**		**1 5**		
200/230 µm	**M M**		**1 2**		

Figure 6.10-5. Fiber attenuation versus light wavelength characteristics. Attenuation per unit length continues to be reduced steadily, through improved fiber drawing techniques and reduction in impurities. It is now approaching the theoretical limits of silica-based glass at the 1300 and 1500 nm wavelengths.

Medium	Refraction index	Propagation time (ns/m)
Vacuum	1.0	3.333
Air	1.003 (rounded to 1)	3.334
Water	1.33	4.444
Fused quartz	1.46	4.860
Glass	1.5	5.014
Diamond	2.0	6.686
Silicon	3.4	11.366
Gallium Arsenide	3.6	12.034

Figure 6.10-6. Typical indices of refraction. Although the index is measurably affected by light wavelength, the influence of wavelength on the index in optical fibers is small enough to be ignored in determining them. *(Table created by integrating copyright information from AMP, Inc. and Force, Inc. publications, with their permission.)*

Numerical aperture (NA) is a measure of the amount of light source output coupled into the fiber:

$$NA = (n_1^2 - n_2^2)^{1/2}$$

Note that the NA's magnitude is determined by the magnitude of the difference of two indices of refraction—the *core* index and the *cladding* index.

The figurative circular *acceptance cone* (ϕ) of the fiber (Figure 6.10-9) is derived from the numerical aperture:

$$\phi = \sin^{-1} NA$$

Larger cones gather more light from rays entering at larger angles, but suffer from higher modal dispersion. The converse is true for smaller acceptance cones. Lower NAs generally translate to a higher *fiber bandwidth.* (The optimum NA for long distance transmission of broadband television signals is about 0.2.)

Optical Fiber Propagation Fundamentals

Fiber Construction Basics

An optical fiber has three concentric components, the *core, cladding* and *buffer* (see Figure 6.10-10). The innermost core is the light conductor. It accepts the light source rays directed into it by the light source lens. Since these rays were independently diverging as they entered the lens, they will converge into the fiber at all angles within an acceptance cone (see Figure 6.10-8) determined by the critical angle (see Figure 6.10-6). The surrounding cladding reflects light impacting its boundary at greater than the critical angle back into the core.

The buffer is the outer cylinder of light absorbing material surrounding the cladding. It may be a solid material bonded to the cladding (tight buffer), or a larger diameter opaque tube in which the fiber runs without tension (loose buffer). The interior of a loose buffer may be filled with an inert gas.

Fiber Types

Of the many ways to classify fibers, the most informative is by refractive index profile and number of modes accepted. The two main types of multimode fiber index profiles are *step* and *graded.* In a step index multimode fiber, the core has a uniform index with a step index change at its boundary with the cladding (see Figure 6.10-11). The lowest order mode travels down the center while higher order modes strike the core-cladding interface at angles above the critical angle. As a result, simultaneously departing angular rays progressively fall behind the first order ray as they propagate down the fiber. This spreading is called *modal dispersion.*

In a graded index multimode fiber (Figure 6.10-11) the core's index is not uniform; it is highest at the center and decreases until it reaches the cladding. This reduces modal dispersion, because the core is essentially a series of concentric rings with increasingly lower refractive indices. Since light travels faster in a lower index medium, light farther out from the fiber axis travels faster. Further, since high order modes have a faster average velocity than low order modes, all modes tend to arrive at a point at nearly the same time. Rays of light are not sharply reflected by the core-cladding interface, they are refracted successively by differing layers in the core.

Single-mode step index fiber has a very small diameter step index core, but has the same cladding diameter as the two most widely used multi-mode fibers (see Figure 6.10-11). Modal dispersion is eliminated by making the core small, typically 7 to 10 μ (1/6 the diameter of a human hair), rather than in excess of 50 μ. At this diameter, only the direct (parallel to the

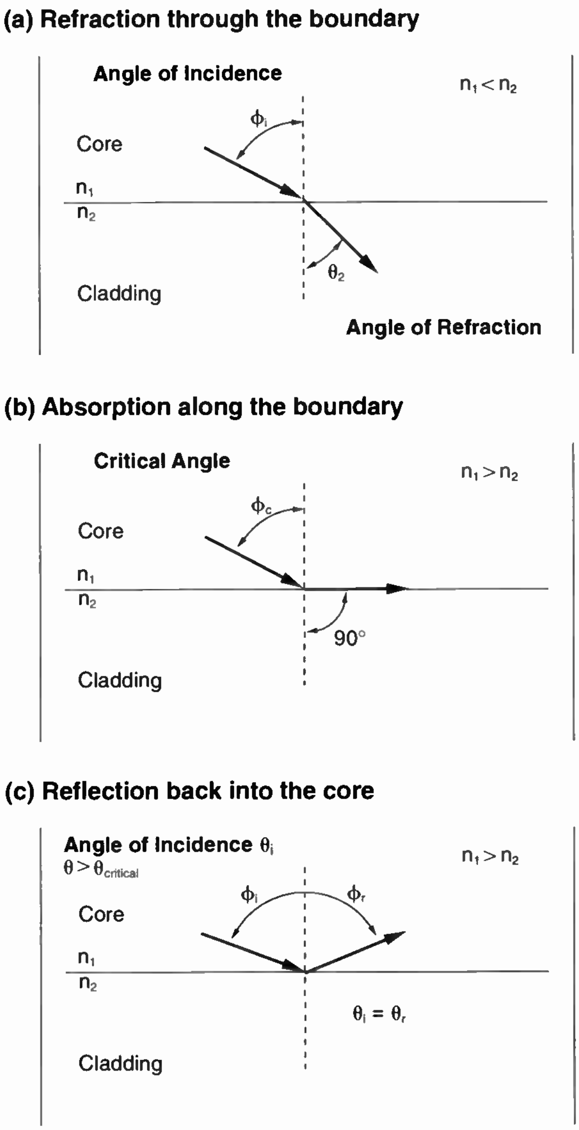


Figure 6.10-7. Lightwave refraction principles. The refraction index of the core n_1 is always less than that of the cladding n_2. Light incident on the boundary at less than the critical angle ϕ_i propagates through the boundary, but is refracted away from the normal to the boundary (a). At the critical angel ϕ_c, along the boundary (b). Light incident on the boundary at angles ϕ_i above the critical angle, is totally internally reflected (c). *(Adapted from Force, Inc. illustration with permission.)*

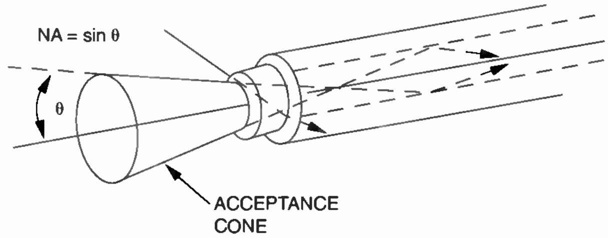


Figure 6.10-9. Light ray acceptance cone geometry. The acceptance cone is an imaginary right angle cone extending outward coaxially from the fiber's core. It is a measure of the light-gathering capability of a fiber. Its ray acceptance angle, called the Numerical Aperture (NA) of the fiber is uniquely determined by the refractive indices of that fiber's core and cladding. *(AMP, Inc. copyright illustration, used by permission.)*

fiber axis) mode is propagated. The single-mode of propagation is appropriate for long distance transmission of high bandwidth signals.

Refraction Index Ratio and Numerical Aperture Relationships

To maximize reflection of bounce-propagating rays, the cladding's index of refraction must be measurably lower than the typical core value of approximately 1.5 (compared to air's 1.003) (see Figure 6.10-12). Decreasing the cladding's index toward 1.40 dramatically increases the fiber's NA, therefore widening the acceptance cone angle and increasing the allowable input power level. On the other hand, this will cause the signal exiting the fiber to have substantially increased modal dispersion.

Fiber Choices

Three designs of optical fiber are now used in the telecommunications industry for broadband signal transmission (see Figure 6.10-13). It is difficult to identify a single best choice, when all performance parameters of all design options are considered.

Multi-mode 50 μm core fiber was the first commercially available in the 1980s. It is no longer installed in common carrier networks, but is popular for private installations because it is easy to terminate and its bandwidth distance product is adequate for some short distance broadband signal transport needs.

Multi-mode 62.5 μm core fiber was originally developed by AT&T because there was less need for precision termination, it had less attenuation, and it would

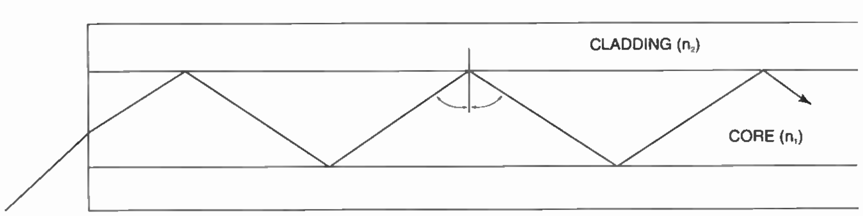


Figure 6.10-8. Total internal reflection. Rays of light incident on the core/cladding boundary at greater than the critical angel, determined by the quotient n_1/n_2, propagate down the fiber's core at a velocity determined by that fiber's value. (See Figures 6.10-6 and 6.10-7.) One ray is shown to keep the diagram simple. (See Figures 6.10-11 and 6.10-12.) *(AMP, Inc. copyright illustration, used by permission.)*

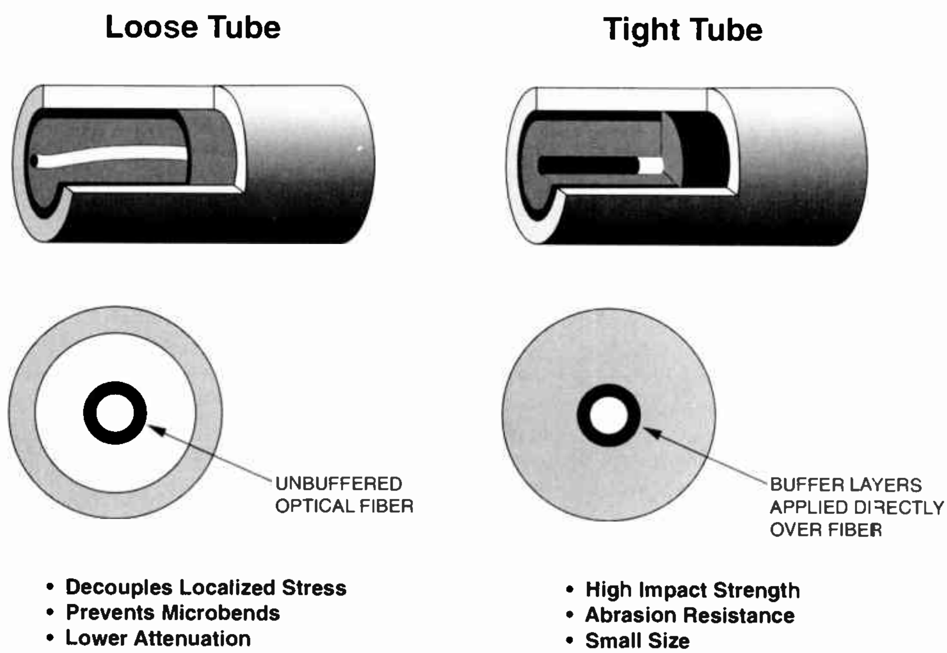


Figure 6.10.10. Optical fiber tube construction. Choice of loose or tight buffer is determined by the physical environment in which the fiber is to be installed. *(Force, Inc. copyright illustration, used by permission.)*

accept substantially higher light power inputs than would 50 μm fiber. These attributes somewhat offset its 50% lower bandwidth-distance product.

Single-mode fiber became the medium of choice for telecommunications industry use when it became widely available in the late 1980s. Its nominal 9 μm core eliminates distance limitations imposed by modal dispersion, the narrower spectral width lasers which it accommodates reduce chromatic dispersion and the order of magnitude bandwidth distance product increase (with respect to 50 μm fiber) far offsets the increased precision needed for termination.

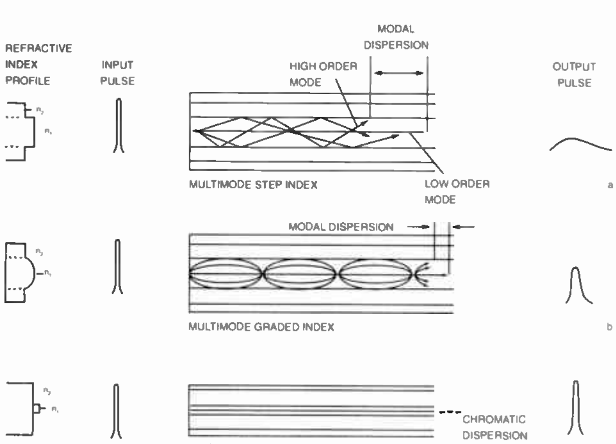


Figure 6.10-11. Optical fiber types. The core diameter and its refractive index characteristics determine the light propagation path or paths within the fiber's core. *(AMP, Inc. copyright illustration, used by permission.)*

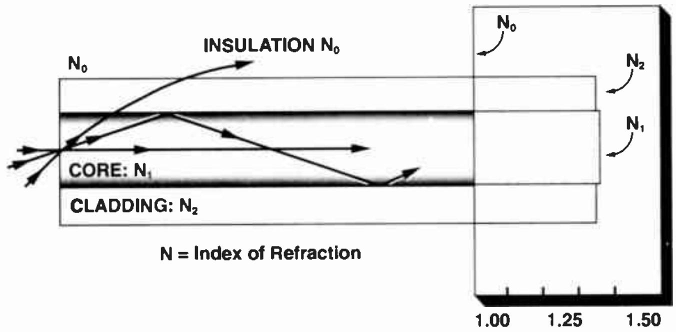
Attenuation

Loss of power (attenuation) during light propagation in a fiber is due to: light absorption resulting from impurities in the core and propagation paths in the cladding and scattering (reflection in the core) caused by impurities in the light path. Fiber attenuation from both causes is specified in decibels per kilometer. For commercially available fibers, attenuation ranges from under 0.5 dB/km for premium single-mode fibers to 1000 dB/km for large core plastic fibers. (Since emitted light is expressed as a power quantity, ±3 dB as a power quantity represents a doubling or halving of any reference power level.) Refer to fiber vendors' catalogs exact absorption and scattering loss figures for specific products.

Attenuation and light wavelength are also uniquely related in fiber transmission systems (see Figure 6.10-5). Most fibers have medium loss values (3 to 5 dB/km) in the 800–900 nm wavelength range, low loss values (0.6–1.5 dB/km) in the 1,150–1,350 nm range, and very low loss values (less than 0.5 dB/km) in the 1,550 nm range. (These wavelength ranges are known as Windows I, II and III, respectively). Optimum fiber circuit performance is achieved by choosing the fiber type and light source wavelength as a function of required bandwidth and transmission distance.

Dispersion

Dispersion in fiber transmission systems produces effects analogous to pulse flattening caused by high frequency attenuation in copper cable, as illustrated in Figure 6.10-14. It is the limiting factor in determining the bandwidth of a fiber circuit, and is most easily



n_2	MATERIAL NA	ACCEPTANCE CONE HALF ANGLE	ACCEPTANCE CONE FULL ANGLE
1.40	.538	32.6	65.2
1.41	.512	30.8	61.6
1.42	.483	28.9	57.8
1.43	.453	26.9	53.8
1.44	.420	24.8	49.7
1.45	.384	22.6	45.2
1.46	.344	20.1	40.2
1.47	.298	17.4	34.7
1.48	.244	14.1	28.3
1.49	.173	9.9	19.9

Figure 6.10-12. Cladding reflective index n_2 influences on NA and acceptance angle of step index MultiMode fiber. The core refractive index n_1 is specified at 1.5. Insulating material (N_0) surrounding cladding absorbs light refracted through the cladding/insulation boundary. *(Illustration created by integrating copyright information from AMP, Inc. and Force, inc. publications, with their permission.)*

understood by examining the propagation of a precise square wave pulse train at the fiber input. As this pulse train propagates down the fiber, each pulse will first become wider and more round-shouldered. Eventually, all the pulses will flatten down to a series of wavering crests indistinguishable from each other.

Modal dispersion in multi-mode fiber is quantified by the spread of ray arrival times, caused by the different path lengths traveled by bouncing rays (as depicted in Figure 6.10-11). These rays may individually be at different wavelengths, but this is of little consequence in multi-mode fiber transmission.

Chromatic dispersion in single-mode fiber is due to the fact that solid-state light sources are not *mono chromatic* (do not emit light of a single wavelength). As previously mentioned, the propagation velocity of light varies directly with its wavelength. The magnitude of chromatic dispersion is a function of both the light source's *spectral bandwidth* and the circuit length. A special class of long wavelength distributed feedback (DFB) lasers with a narrow output spectrum is now available for propagation of information on single mode fibers over 50 km or more.

Bandwidth-Distance Products

Dispersion effects, and length of a given fiber transmission circuit combine to uniquely define the bandwidth or information carrying capacity of a fiber. This situation has no equivalent analog in electrical signal transmission theory. The quantification of this performance parameter is a multiplication product called the *bandwidth-distance product,* expressed in megahertz/kilometers, and serves as a numerical figure of merit for a fiber (see Figure 6.10-13).

Again, using the propagation of a pulse as an example is helpful. Fourier theory establishes that a precisely square cornered pulse consists of a fundamental frequency plus an infinite series of in-phase odd harmonics. Square wave pulse train flattening in fiber transmission, caused by modal or chromatic dispersion, is analagous to pulse train flattening in copper transmission caused by circuit capacitance.

High rate data streams are essentially streams of short pulses. By the bandwidth distance concept, it is obvious that for a given fiber, bit rates must be slower when the data stream must travel longer distances. When the direct rays representing the leading edge of bit "n" in a stream arrive before the last reflected rays representing the trailing edge of bit "n − 1," bit errors will occur and the delivered signal is likely to contain little or no recoverable information. Dispersion phenomena must therefore be considered when choosing a fiber type, so that the bit stream to be transmitted over a given circuit length will not arrive with unrecoverable bit errors.

In choosing a fiber type for installation in a given transmission circuit, the product of the bandwidth (in MHz) to be transmitted and the length of the optical

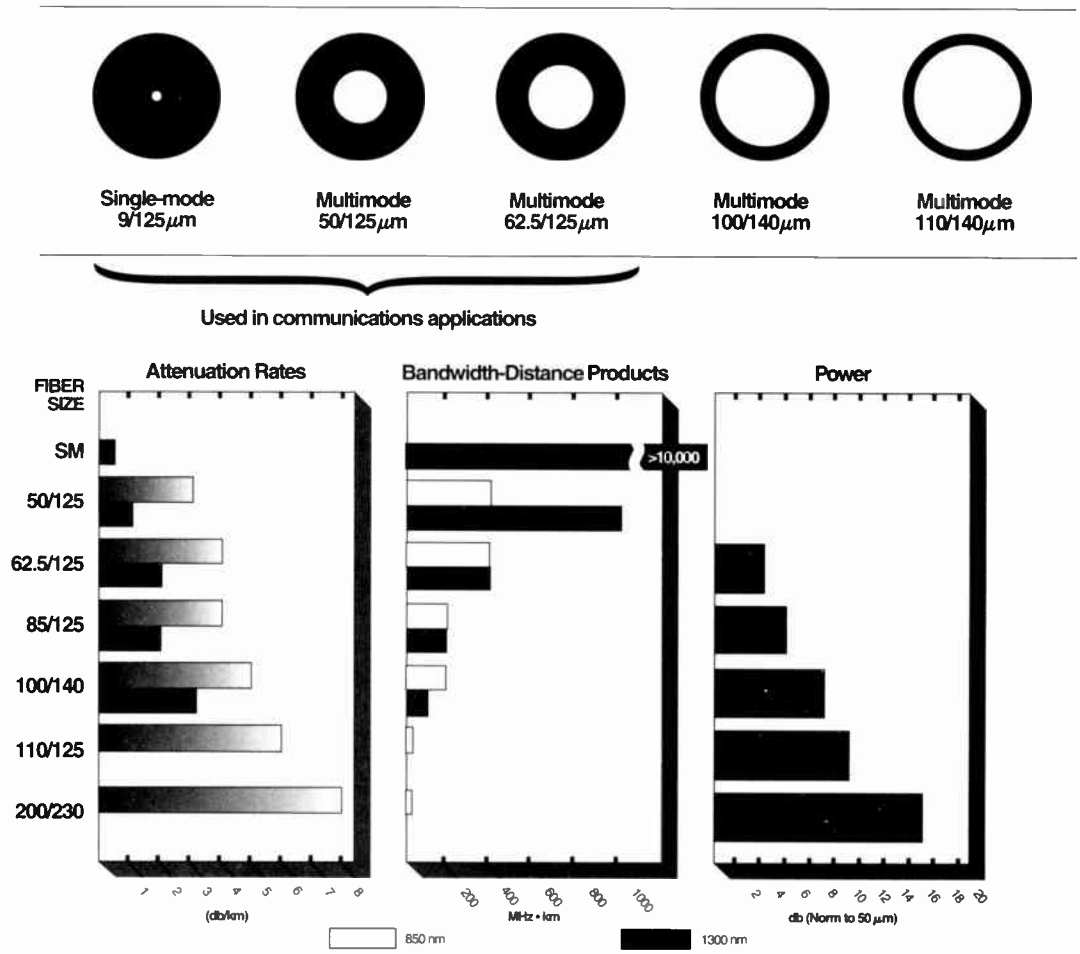


Figure 6.10-13. Popular fiber sizes. Attenuation per unit length, bandwidth distance products, and power handling capabilities are determined by the magnitude of the core diameter and the core's material and refractive index. *(Illustration created by combining Two Force, Inc. copyright illustrations, used by permission.)*

circuit (in km) must be less than the bandwidth-distance product figure of merit number (see Figure 6.10-13).

Fiber Transmission System Components

Optical sources in fiber communications transmission systems are either *light emitting diodes* (LEDs) or *injection laser diodes* (ILDs). The LED is a more incoherent source (it has a larger spectral bandwidth) than is an ILD. The LED is characterized by output power levels well below 0 dBm in the order of -10 to -30 dBm. It operates at slower speeds, because of the hysteresis (lag) of changes in light intensity to changes in input voltage (see Figure 6.10-4). However, its cost is considerably less than that of an ILD. It is quite suitable for applications requiring transmission across distances under 10 km and modulation bandwidths that accommodate data streams of less than 100 Mbps. LEDs operate in the 850 and 1300 nm windows.

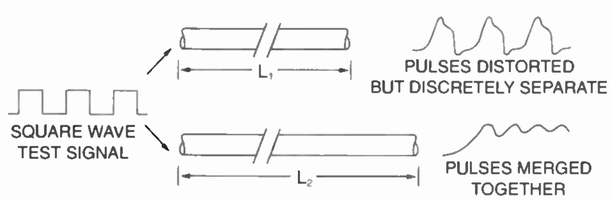


Figure 6.10-14. Lightwave pulse propagation principles. Modal and chromatic dispersion characteristics of a fiber determine the maximum frequency of modulating signal which can be recovered at the end of a fiber of any given length.

ILDs are capable of power outputs on the order of 0 dBm and higher. Upper power limits are a matter of device cost and permissible radiation levels. Although ILDs cost substantially more than LEDs, they operate in the 1300 and low attenuation 1550 nm windows. Their combination of low attenuation and high modulation frequency make them cost justifiable as sources in systems which must operate over extended distances, or which must transport large numbers of signals or both.

Selecting an optical source for a fiber transmission system requires a careful evaluation to insure that its modulation frequency limit is greater than the bandwidth to be transmitted, and that it provides enough optical power at the appropriate wavelength, to satisfy the distance requirement. Another important consideration is the fact that optical devices do not turn on with the linear characteristics of solid-state high speed gates (see Figure 6.10-4).

Light source cost factors are constantly changing, driven both by cost/demand volume relationships and manufacturing process refinements. Historically, low cost 850 nm devices were adequate for short distance transportation of narrowband signals. Today, 1,300 nm LED sources are popular because of their unit costs and lower attenuation figures. However, 1,300 nm lasers driving single-mode fiber have been standardized as components of choice in two SMPTE Standards—*SMPTE 292 M for 1.485 Gbps HDTV Signal Transmission* and *SMPTE 297 M for 143 Through 360 Mbps Digitized SDTV Signal Transmission.*

Both the telecommunications and the television industry are making increasing use of 1550 nm DFB lasers and single-mode fibers, both because laser prices are beginning to fall and dispersion-shifted fiber is becoming available.

Light Detectors

Light detectors perform a complementary function to light sources—converting incident optical energy to electrical energy. Detectors in fiber transmission systems are positive intrinsic negative (PIN) or avalanche photo diode (APD) semiconductors. The limiting sensitivity (threshold) in detecting weak incoming signals determines link performance.

In an ideal PIN diode, each incident photon creates an electron hole pair in the semiconductor lattice. This in turn sets one electron flowing in the external circuit. If the received light is weak, the generated current may not be strong enough to overcome the noise inherent in the diode and receiver circuit. In such cases, it is desirable to increase the detector output before processing by the receiver. Such amplification is generated in an avalanche photo diode. A reverse bias adds several electron volts of energy to each liberated electron to provide amplification of the received signal.

Repeaters

A fiber transmission circuit's maximum useful length is a function of six independent variables:

- Signal bandwidth to be transported
- Transmitter launch power
- Fiber, splice and connector pair attenuation
- Modal and chromatic dispersion characteristics
- Receiver threshold sensitivity
- Minimum end-to-end transmission circuit performance specifications set by the user.

Analog circuits 50 km long are now possible. Digital circuits equipped with bit error detection and correction overhead can operate for perhaps this distance and more.

The first generations of analog and digital repeaters developed to extend these transmission distances operated at electrical baseband. The repeater terminal is literally a receiver whose demodulator is hardwired to the modulator of a transmitter. This design is crudely equivalent to creating a long distance telephone circuit by holding the earpiece of one old-fashioned two-handed telephone to the mouthpiece of another.

Analog domain repetition in this fashion has a limit of four or five tandem sections and a transmission distance maximum of 150–200 km. Digital domain repeaters (regenerators) have no such limit, as long as bit errors are detected and corrected at each regenerator prior to retransmission.

Second generation repeaters operating entirely in the light domain have no direct equivalent in electrical signal transmission products (see Figure 6.10-15). Two signals are input to the passive coupler (an 8 m length of Erbium-doped fiber)—the 1,300 nm optical signal to be amplified and a 10–50 mW pump laser operating at 980 or 1490 nm. The photonic interaction of the pump laser with the optical carrier signal results in the synthesis of a new optical carrier with 20–30 dB higher power, modulated with the original information, but operating at a nominal 1550 nm wavelength.

This increase in wavelength (from 1,300–1,550 nm) presented a significant problem to longhaul common carriers with fiber networks optimized for 1,300 nm carrier transmission. Further research and development has resulted in equipment with amplified output carriers operating within the 1,300 nm window.

Switching

First generation fiber transmission circuit switchers operate at electrical baseband. Advantages of optical transmission and switching are not confined to broadband video image transmission, fortunately. Optical domain switching and interconnection of computer components is necessary to develop the next generation of super high speed, massively parallel number crunching engines. Efforts to develop optical switchers are underway in telecommunications and computer industry advance research labs.

First generation all-optical switchers will reach the market early in the next decade.

Video and Audio Signal Transmission Fundamentals

Information to be transported by optical fiber transmission systems may be modulated as amplitude mod-

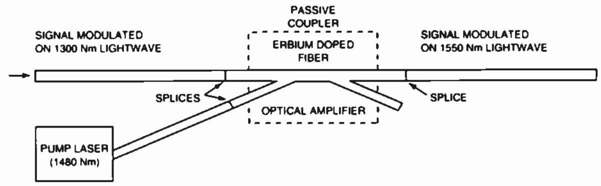


Figure 6.10-15. Block diagram of first generation optical amplifier (repeater).

ulation (AM), frequency modulation (FM) or digital pulse code modulation (PCM) waveforms.

Amplitude Modulation

AM is the simplest form of modulation. The instantaneous amplitude of the input signal directly controls the output intensity of the light source. For an NTSC standard video input signal, the *peak white* maximum amplitude results in maximum light intensity and the *tip-of-sync* the minimum. Because of the nonlinear transfer characteristic (see Figure 6.10-4) of LED sources, intermodulation distortion (IMD) is introduced into the fiber-borne video signal when such modulation characteristics are implemented. In addition, the Signal to Noise Ratio (SNR) of the signal is directly correlated to the optical power, and attenuation in a system lowers SNR directly. Further, since most optical devices do not completely *turn off* (particularly lasers) there is always an optical noise floor present. AM modulation technique is no longer used in the design of video broadcast signal transmission products.

Frequency Modulation

FM overcomes the SNR and IMD limitations of amplitude (intensity) modulation. FM requires more electronics, but removes the problems of signal quality from the optical domain to the electrical domain. For FM, ancillary audio and data signals are first multiplexed over the video baseband onto separate FM subcarriers. This video/audio signal group is then frequency modulated on a carrier in the 30 MHz range (and higher), to produce a single signal with a fixed amplitude excursion. Its value then modulates the light source over a linear portion of its range.

Required SNR and IMD performance specifications are achieved in the design of the frequency modulator.

Digital PCM

A fiber transmission system is most transparent when it is operated in a pulsed mode. It is therefore particularly complementary to signals which are presented to it as fixed amplitude FM or a digital bit stream. In telephony applications, the relatively narrow bandwidth of a single voice channel can be efficiently digitized and multiplexed with up to 10,000 + other similar channels on a single fiber. Received quality of a digitally transmitted voice channel is markedly better than analog, because of the complete absence of thermal noise, impulse noise and distortion endemic to analog transmission.

Electrical Driver Signal Multiplexing Techniques

Frequency Division Multiplexing (FDM)

The FDM *technique of combining multiple AM or FM* carriers is widely used in coaxial cable distribution based systems. Unfortunately, the nonlinearity of optical devices operated in the intensity modulation mode results in substantial and often unacceptable noise and intermodulation distortion in the delivered signal channel. Wide and selective spacing of carriers ameliorates this problem to some degree. Some FDM fiber transmission system specifications, although unacceptable to broadcasters for backhaul applications, nevertheless represent a gratifying improvement in video (and audio) signal quality delivered to cable system subscribers or transmitted in videoconferencing networks.

Digital Time Division Multiplexing (TDM)

Television signal group transmission systems of the 21st century will operate entirely in the digital domain. All the technology needed to create, process, store and transmit images in the digital domain has already been invented. The issues of when, why and how the day of all-digital operation will arrive are currently in the realms of politics and economics. When resolved, they will establish an environment which fosters and drives fast issuance of many needed standards.

Wavelength Division Multiplexing (WDM)

Multiplexing optical streams in the optical domain reduces the number of optical fibers required to meet any specific transmission requirement (see Figure 6.10-16). Multiple independent fiber transmission systems operating at different optical wavelengths can be transported over a single fiber as is, by combining them in a passive optical multiplexer. This is a simple product in which optical feeds from multiple optical transmitters are fused together and then spliced into the transporting fiber. Demultiplexing these optical signals at the receiver end of the circuit is accomplished in an opposite oriented passive optical demultiplexer. The multiple output fibers are coupled into photodetectors through wavelength selective optical filters.

Options for coupling light sources of different wavelengths were very limited when the *NAB Engineering Handbook, Eighth Edition*, was written. In the most rudimentary design, a fiber could transmit two or three lightwave circuits without receiver optical filtering, one wavelength each from the 850, 1,300 and 1,550 nm windows, traveling in the same or opposite directions. A more advanced feasibility demonstration system used 1,280, 1,300 and 1,320 nm wavelengths from Window II to transport an analog HDTV video signal's red, green and blue components. Sharply tuned optical filters provided received signal separation. Audio signals were multiplexed on a separate optical carrier.

Since then, AT&T Bell Labs and numerous other fiber R&D labs have been focusing on development of *dense wave division multiplexing* (DWDM). The design objective was to increase fiber circuit capacity and revenues sufficiently to offset costs of the DWDM multiplexers. One early 1997 Bell Labs demonstration featured a single semiconductor laser whose light output could be filtered into over 200 discrete light sources. Laser manufacturers can now offer a range of products operating at guaranteed specified wavelengths within the window, with narrow spectral bandwidths in the order of a few nanometers. Optical filter manufacturers similarly are now offering products with

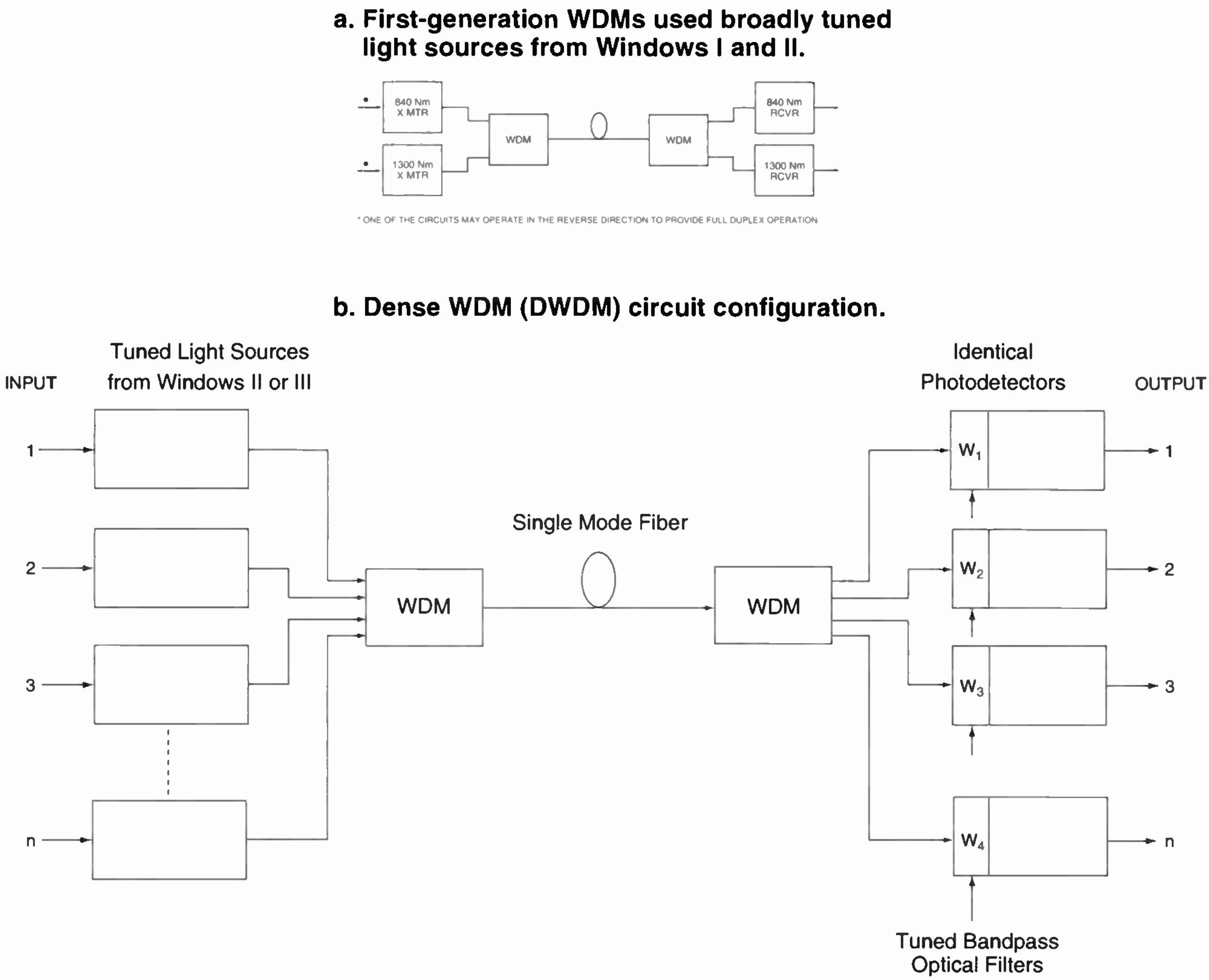


Figure 6.10-16. Wave Division Multiplexer (WDM) circuit configurations. WDMs are passive devices created by fusing optical fiber pigtails (short lengths) together. The most uniform of a fiber circuit transporting multiple light wave carriers results from use of specifically tuned light sources from either the 1300 or 1550 nm Windows II or III (DWDM), one single-mode fiber may be used for full duplex (FDX) operation by transmitting the Window II and III carrier bundles in opposite directions.

narrow passbands centered around specific wavelengths.

Many new considerations are therefore important in specifying DWDM systems. First, adequate separation (measured in dB of rejection) between optical channels is required. Optical crosstalk between channels results in lowering the effective SNR of the link.

Secondly, installation of passive multiplexers and optical filters in fiber circuits results in significant optical attenuation (1–5 dB per device). As a result, the maximum transmission distance is reduced.

Lastly, the output wavelength of optical light sources can change over temperature. Consequently, the tuned bandwidth of the WDM components must be wide enough to accommodate this drift.

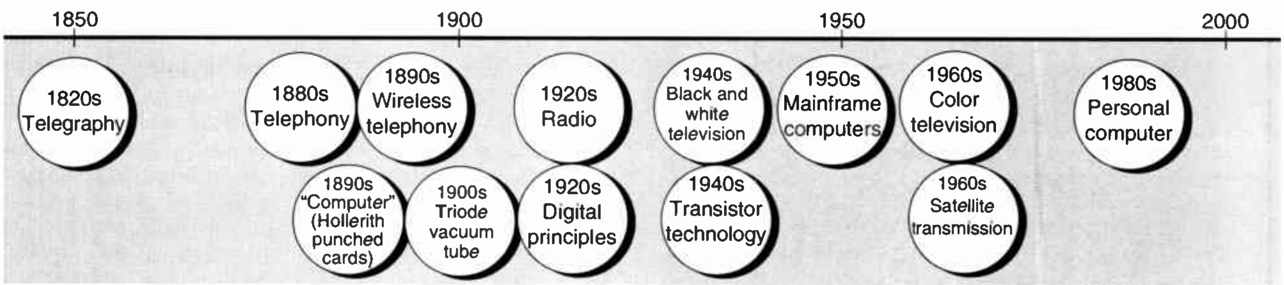
Constantly changing cost/benefit equation variables and their assumed constants are making WDM and DWDM more cost justifiable. However, costs for defining and achieving adequate, economically justifiable optical isolation are unknown. Further, the techniques are only appropriate for digital systems whose carrier to noise characteristics can overcome power losses in the WDM components.

SYSTEM DESIGN AND APPLICATIONS EVOLUTION

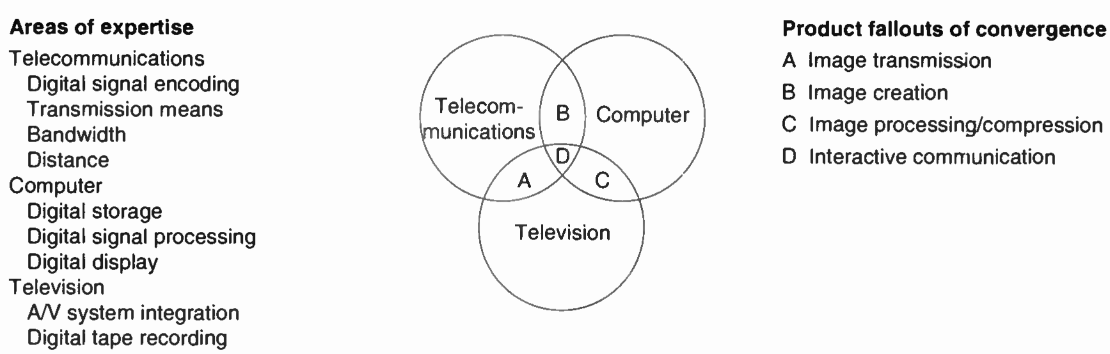
Applications of telecommunications, computer and radio/television technologies began as distinct and separate solutions to common human communication needs (see Figure 6.10-17(a)). They were designed to provide reception at a distance of information, entertainment and/or education, as aural or visual inputs to the ears and eyes. During WW II, however, computer technologists were stifled by the performance shortcomings in the vacuum tube technology developed to process and transport electrical analog signal forms.

AT&T Bell Labs researchers Shockley, Brattain and Bardeen provided the needed technological breakthrough in 1947—transistor technology. Embracing it as their own undergirding technology base, computer technologists in the late 1960s began to develop digital signal processing products whose speed and versatility and cost and size immediately began to follow Moore's Law predictions (see Figure 6.10-18). And thus began a period of inexorable convergence of these once disparate radio/television, computer and telecommunications technologies (see Figure 6.10-17(b)).

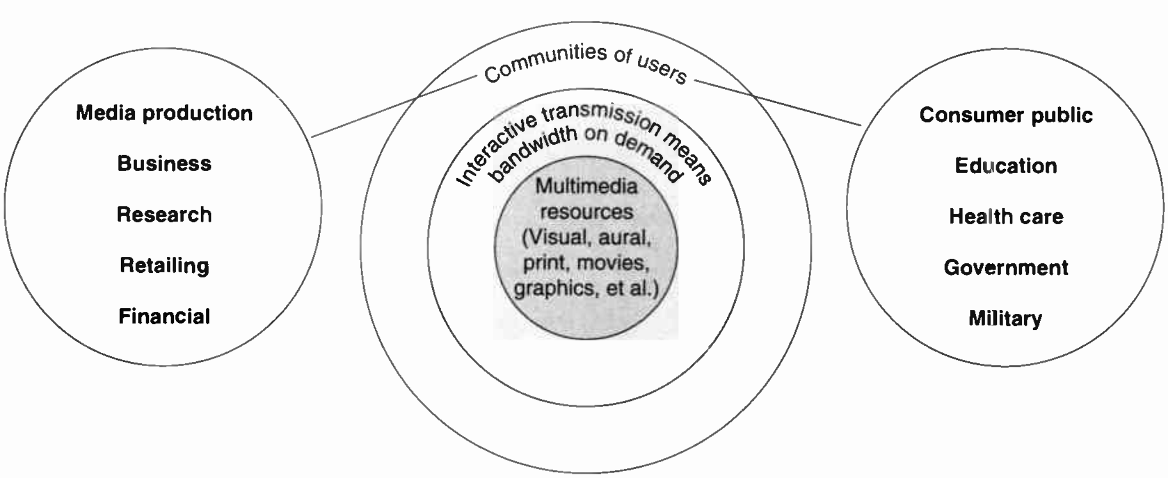
Convergence of these technologies will continue until they are one and they foster the development of one universal, all-digital, two-way interactive entertainment/information/education system (see Figure 6.10-17(c)).



a. Time line of disparate communications technology births.



b. 1960s beginnings of disparate technologies convergence.



c. 21st century spheres of electronic communications production, storage and dissemination.

Figure 6.10-17. Technology development relationships. Over the past 150 years they have changed from disparate and totally unrelated (a) to overlapping and converging today (b). If convergence overlapping continues into the 21st century the result might be one technology creating products for all information transportation applications (c). That end requires that many globally dispersed organizations issuing interface standards all agree on one standard, an unlikely occurrence.

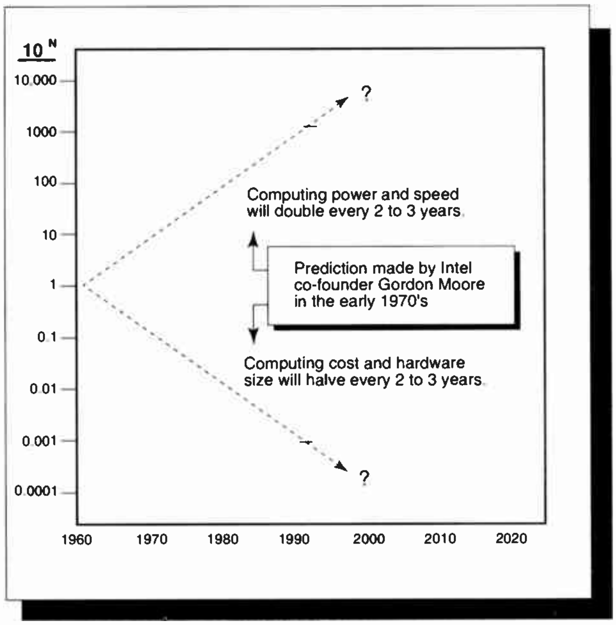


Figure 6.10-18. Semi-log plot of *Moore's Law.* Dr. Gordon Moore postulated it in the 1970s a few years after he co-founded Intel, Inc. Based on Intel's early years of chip performance/pricing trends, the Law still generally applies. The 2 to 3 years doubling/ halving period has shrunk to about 18 months, and may become less than a year in the 21st century, and has no predict able end. Applying the Law to the manufacturer of automobiles, its import becomes astounding. A Mercedes or Cadillac which could be purchased in the 1960s for around $10,000, provided about ten miles of travel per gallon of gas. Their tenth generation successors purchase in the 1990s would have been priced at under $10.00, and would have provided 10,000 miles of travel per gallon!

All-digital Television Operations

This section's purpose is to highlight the emerging indispensable role of fiber optic technology in designing future proof 21st century television signal transmission systems. This technology must initially coexist with but ultimately replace many of the copper wire and microwave wireless systems which have served the industry's 20th century signal transportation needs.

Broadband/Wideband Considerations

To telecommunications engineers, *broadband* describes a transmission circuit with a throughput of about 1.5 Mbps maximum. The 45 Mbps (DS-3) digital transmission scheme for 4.2 MHz NTSC video signal transmission is *super broadband.* This service has been marketed to broadcasters with the same finesse used by Henry Ford in the early 1900's: "We'll sell you any color car you want as long as its black." As television producers and broadcasters embrace each new bit rate (serial digital NTSC, "601" component video, MPEG-2 MP/ML, MPEG-2 4:2:2P/HL for HDTV, etc.), a new interface at that transmission rate is eventually introduced by a pioneering codec manufacturer. However, the bit stream is diligently reduced to a payload rate below 45 Mbps before it sees the light.

To television production and broadcasting engineers, broadband describes a transmission circuit which will accept an uncompressed video signal with per color analog bandwidth as high as 8–10 MHz for NTSC R, G and B and 20–30 MHz for HDTV. It also describes the digital equivalent throughputs of 270 Mbps (digital component 4:3 video and four AES/EBU audios, per SMPTE Standard 259M) and 1.485 Gbps (digital component 16:9 video and eight AES/EBU audios, per SMPTE HDTV Standard 292M). Those signals are the real numbers for broadcast quality.

To top end non-broadcast television users (educational and medical applications), a 1.544 Mbps (T1) transmission circuit is broadcast quality (better than S-VHS), most acceptable for transmission of full-motion NTSC color video (143 Mbps uncompressed). Corporate communications videoconferencing and retailing industry users accept sub-T1 circuits as broadcast quality.

Transmission Circuit Location Designators

Popular telecommunications industry acronymic designators of circuits used for signal transmission are local, metropolitan and wide area networks (LANs, MANs and WANs). These three descriptors are inadequate to describe the spatial areas which TV producers and broadcasters must traverse with transmission circuits—buildings, campuses and oceans. Figure 6.10-19 summarizes these (in italics to denote they are the author's creations), along with descriptions of their spatial coverage areas.

Fiber Technology History and Applications Evolution

Principles of lightwave transmission by total internal reflection in a stream of water were first demonstrated to London's Royal Society in 1870 by Sir John Tyndall. In 1880, Alexander Graham Bell used modulated sunlight to transmit voice communications in his patented photophone. During the pre-WW II years, researchers developed systems for transmitting images through glass rods.

After the WW II hiatus of all technological development for commercial purposes, practical applications of the principle of total internal reflection concentrated on glass rods. Development of the medical fiberscope in the mid 1950s prompted glass rod researcher N. S. Kapany to name the technology fiber optics. In 1958, Bell Labs scientists A. L. Schawlow and C. H. Townes were the first to convert electrical energy to light energy in a device they named the *laser* (light amplification by the stimulated emission of radiation). Lasing action in a Gallium Arsenide (GaAs) semiconductor was achieved in 1962.

Research to develop commercially viable applications of lightwave technology began immediately (see Figure 6.10-20). In the 1970s, the telecommunications industry began to see the light through installations of first generation fiber optic transmission systems. In the 1980s, fiber cables began to replace terrestrial and satellite microwave at a rapid pace, for continental and

Acronym	Descriptor	Location
LAN	Local	Within a room or part of a building, generally installed and owned by the premises tenant (UTP and coax cables)
BAN	Building	Circuits connecting LANs within a building, which may be installed and owned either by the premises tenant or the building's landlord (UTP, coax and MM fiber cables)
CAN	Campus	Circuits connecting BANs within a user's premises, both installed and owned by a local common carrier and bought by the premises owner for installation by an outside contractor (UTP, coax, MM and SM fiber cables)
MAN	Metropolitan	(Generally) Circuits installed on, under or over public rights of way, installed and owned by LECs (Local exchange carriers) or ASPs (Alternative service providers) (SM fiber cables and terrestrial pt-pt microwave)
WAN	Wide	Large groups of circuits interconnecting MANs within a contiguous land mass, generally also within a country's political boundaries, installed and owned by IECs (Inter-exchange carriers) (SM fiber cables and satellite microwave)
GAN	Global	Large groups of circuits interconnecting WANs, both across countries' political borders and across ocean floors (SM fiber cables and satellite microwave)

Figure 6.10-19. Hierarchy of circuit location and function designators.

transoceanic long distance telephone call transmission, respectively.

By 1990, lightwave conducting fiber cables had become the backbones of the world's long distance telephone networks. Multiple fiber cables carrying upwards of 10,000 digitized voice and computer digital data channels on each fiber were strung on telephone poles, buried in conduits and trenched under ocean floors. Fiber installation in intracity trunking cables is now a fast-growing application. Cable industry use of fiber in interhead end backbones and fiber-to-the-curb (FTTC) is similarly fast growing.

Fiber Transmission Advantages

Advantages of fiber optic transmission over alternative copper wire and terrestrial and satellite microwave media have not changed, except in magnitude, since they were first articulated in the 1980s. These advantages include the following:

- **Bandwidth.** Maximum information carrying bandwidth of a transmission system is determined by the frequency of the carrier on which the information is modulated. The 30–300 THz spectrum of infrared light frequency carriers are three to four orders of magnitude (1,000–10,000 times) higher than the 1.5 Gbps highest digital video frequency found in an HDTV production facility (see Figure 6.10-3).

  The enormous transmission bandwidth capabilities of lightwave systems are expressed in megahertz-kilometers. Single-mode fibers have bandwidth distance products of 10,000 and higher (see Figure 6.10-13). That translates into a 10 GHz bandwidth for a 1 km fiber circuit. At a transmission packing density of 4 bits per cycle, the digital throughput maximum of this circuit would be 40 Gbps.

- **Low loss.** Fibers provide substantially lower attenuation than copper cables and twisted pairs and require no equalization circuits to correct for high frequency rolloff. Signal attenuation in premium quality single-mode fiber is of the order of 0.5 dB/km and is falling as this technology is improved. Unrepeatered transmission of high bit rate digital signals over 50 km and longer is now commonplace. Signal loss in splices, connectors and patch panels is constantly falling, from 1 dB each downward toward 0.1 dB, as mechanical precision in connectorization and splicing steadily improves.

  The available power of this bandwidth can alternatively be utilized inside a facility. Digital video, audio and data signals can be pushed through the typical maze of mechanically spliced and connectorized tandem cable sections, switch points and patch panels interposed between source and destination equipment.

- **Optical and electrical signal characteristics.** Fibers and their coatings are composed of dielectric

DECADE	ANALOG SYSTEMS	DIGITAL SYSTEMS
1950s	Development of gas laser light sources.	Early research in transmission of digitized sound signals.
1960s	First implementations of fiber optic transmission system principles.	Early efforts to develop multiplexed systems for digitized voice and data transmission.
1970s	First practical systems for analog signal transmission over fiber, through intensity (amplitude) modulation of light source by signal baseband video signal combined with FM subcarriers for ancillary sound and data channels.	Common carrier industry implementation of DS() hierarchy of transmission services for digitized voice and data. First long-haul fiber optic transmission systems.
1980s	1984 - Introduction of single television channel per fiber (STVC/F) transmission systems with baseband video and multiplexed audio FM'd onto nominal 30 MHz carrier. 1984 - Introduction of short-distance multi-fiber RGB graphics systems. 1988 - Introduction of proprietary multiple television channel per fiber (MTVC/F) systems.	Rapid deployment of fiber in common carrier long haul circuits and IEC trunks. Sporadic utilization of DS3 services by broadcast industry. Utilization of sub-DS1 services in videoconferencing. First implementations of STVC/F and MTVC/F systems for full bandwidth NTSC signal transmission.
1990s	Rapid deployment of MTVC/F trunking and distribution networks in local cable systems. Continuing desultory sales of STVC/F systems in private networks. Sporadic broadcast industry utilization of MTVC/F Metropolitan Area Networks (MANs) for back haul of news and local remotes.	First implementations of B-ISDN and SONET systems for transmission of full bandwidth NTSC television signals and high resolution graphics and HDTV signals.
Early 21st Century	Analog transmission systems become historical curiosities.	User setup and use of digital common carrier networks for image transmission become as simple and cheap as DDD phone calls.

Figure 6.10-20. Chronology of development of fiber systems for analog and digital transmission of video, audio and ancillary data signals.

material, which serves to electrically isolate the transmitter and receiver terminals in a fiber circuit from one another. Hum and interference caused by differences in the earth ground potential of separated installations are thereby eliminated.

Complete isolation of separated installations from each others' electrical grounds is assured if the strength material (messenger) in multiple fiber cables is also a dielectric. Also, the *shortest route* between source and destination locations can be chosen, since proximity of fiber cables to equipment with large motors and power transformers does not create problems.

Finally, lightwave transmission produces no electromagnetic radiation which can couple into adjacent fiber or copper circuits. Conversely, lightwave transmissions are impervious to contamination from external sources of electromagnetic interference (EMI) or radio frequency interference (RFI).

- **Size and weight.** Single optical fibers used in television signal transmission have diameters less than a human hair (typically 900 μ—microns millionths of a meter). Copper transmission circuits require a pair of wires, each many times larger, stiffer and heavier than a fiber. A group of cables to carry the same quantity of signals as one fiber will be a hundred or more times larger and heavier than that fiber.

  Total installation costs are much lower for either a temporary field or a permanent studio fiber installation than for a similar copper-based coax or triax system. Fiber cable is the only alternative for circuit capacity expansion when ducts are full of copper, through replacement of each copper circuit with a bundle of fibers. Most common carriers now install fiber on all new construction except for subscriber loops.

  Technicians pulling fiber cables are far less likely to incur workman's compensation covered injuries. Further, they can complete installs and takeouts in 50% or less than the time it takes for coax and triax. This advantage should have the same impact on production truck design as the switch from three camera pickup tubes to three CCD sensors impacted portable and studio camera packaging size and styling.

- **Ruggedness.** All-dielectric cables with appropriate outer sheathings and environment-resistant connectors can be installed in toxic, liquid, flammable and extreme temperature environments and buried without compromising signal transmission integrity. Ruggedized (MIL SPEC tactical) cable designs survive *almost* any hostile treatment during installation or use. Cruelties still to be overcome are those inflicted by installers who playfully bend cables severely to see if they can be broken (they generally can't, but microbends appearing at core/cladding interfaces will impair light flow) and golf course lawnmowers that suck up and twist cables until they both they and the lawnmower are damaged.

  Golf shoe, ski pole spikes and tank treads are not predators which can crush or penetrate MIL SPEC cables. Bending and tension caused by traverses over parapets several stories above ground have no effect on fiber health or performance. In fact, glass fiber has tensile strengths in the order of 600 kpsi, compared to copper's nominal 100 kpsi.

- **Safety.** Lightwaves can not cause spark hazards like electron currents. Careless handling of a fiber cable can not cause bodily harm. Dropping it in a puddle will not damage equipment connected to it.

- **Security.** Lightwave signal flow produces no external manifestation of its existence. Stealing the signal therefore requires physical intrusion into the fiber, which lowers the fiber's output light intensity. Appropriate light output level monitoring makes undetected theft unlikely. Costs for encryption and decryption equipment can thereby be avoided.

Television Industry Applications Chronology

The advantages of fiber over copper for signal transmission quickly became obvious to telecommunications industry systems designers, and motivated a few early 1970s television industry visionaries to test out commercially viable applications in the transmission of electrical signals. Justifications for the employment of fiber continue to proliferate and grow in magnitude.

Tutorial Appendix

TELECOMMUNICATIONS INDUSTRY MAGIC NUMBERS

Analog-to-Digital Signal Conversion Mathematics

Television industry *magic numbers* began being incorporated into SMPTE interface standards in the mid 1980s. Their specific values were not created arbitrarily by television engineers, however. Their synthesis was rooted in the *Nyquist theorem*, first articulated by AT&T Bell Labs scientist and mathematician H. Nyquist in 1924 and 1928.[2] Fellow Bell Labs scientist Claude Shannon defined limits of a noisy channel's Nyquist signal transmission capabilities in 1948.

In brief, 50 to 75 years ago, Nyquist and Shannon wrote the book which today guides engineers developing digital television, radio and audio equipment and transmission media. Two related precepts can be articulated from their bodies of research.[3] Nyquist defined the minimum sampling frequency, for converting analog signals with specified high frequency cutoffs to binary serial digital representations, as twice the highest frequency to be recovered. Setting the sampling frequency of an A/D converter at this frequency will enable reconstitution of the digital bit streams back to exact replications of the original voltage-analog varying signals. Shannon defined the maximum bit rate which can be transported on transmission channels with measurable amounts of Gaussian (continuous, hissy) noise at their outputs, as a logarithmic function of the signal's channel input signal power "S" related to the channel's output noise "N" power. The mathematical expressions for Nyquists and Shannon's theorems are:

$$f_S = 2f_C \text{ (Nyquist's Theorem)} \qquad (1)$$

where f_S is the A/D converter sampling frequency in Hz, and f_C is the highest (cutoff) frequency in Hz of a spectrum of analog signals which can be transmitted over a noiseless channel; and

$$C = W \log_2(1 + S/N) \text{ (Shannon's Theorem)} \qquad (2)$$

[2] H. Nyquist, "Certain Factors Affecting Telegraph Speed," *Transactions A.I.E.E. 1924* and "Certain Topics in Transmission Theory," *Transactions A.I.E.E. 1928.*

[3] Martin, James, "The Maximum Capacity of a Channel," *Telecommunications and the Computer*, Third Edition, Prentice-Hall, Inc., Englewood Cliffs NJ, Chapter 23, p. 413 et seq 1990, 1976.

where "C" is the maximum bit rate (bits/sec) which can be recovered, "W" is the bandwidth (highest analog frequency being digitized and transmitted), "S" is the *rms* power of the transmitted signal measured at the channel input, and N is the *rms* power of the white noise (random, Gaussian fluctuations) measured in the absence of signal at the channel output.

Note that in Equation 1 the Nyquist sampling rate is the minimum needed to digitally represent the entire analog signal bandwidth as a two-state (1, 0) binary digital signal (1 bit resolution), for transmission on noiseless channels. Equation 2 establishes that channel noise disqualifies 1 bit resolution as inadequate to guarantee full bandwidth analog recovery of the transmitted signal.

To hold the recovered signal bandwidth as close to the original cutoff frequency of the sampled signal as possible, SMPTE Standard 259M first specified a minimum resolution (bits per sample) of 8 bits. This resolution provides digital code representations of a 1 V signal for 255 levels (nominal 4 mV) between 0 and 1 V (2^8 bit resolution = 256). This resolution was and still is adequate for digital representation of analog composite NTSC and PAL signals.

However, 10 bit resolution is now used in digital component video signal processing systems. Postproduction house users of the first generation of digital component video editing and recording systems immediately discovered that second generation tape playbacks of editing sessions contained *coffee rings* of hue shifts of saturated colors in flat fields. This was caused by the D/A converter's (DAC) toggling of the eighth bit between 0 and 1 in the presence of white noise approaching 4 mV amplitude.

For system designers' information, the coffee ring syndrome occurs primarily in second generation images created originally by essentially noiseless graphic imaging computer work stations. Television camera outputs contain enough thermal noise to break up the ring pattern into uniform acceptable noise throughout a flat color field. However, concatenation of digital numbers created in computer generated graphics can also create flat fields. This process can result in long strings of binary 1s or 0s, which can cause a DAC to lock for the remainder of each scanning line on which these long strings are present.

This occurrence of this unacceptable color impairment is reduced but not entirely eliminated by increasing sample resolution to 10 bits. The SMPTE standard's specification change was easy to effect. The original issuance provided for unassigned ninth and tenth bits in anticipation of such a need.

The presence of these illegal bit strings can have an even more deleterious impact in a fiber optic transmission system's electrical-to-optical (E/O) converter stage. These unpredictable aberrations seen in a reconstituted analog video image, after its digitization and transportation within a television facility, are therefore easily multiplied when a fragile signal is subsequently digitally reprocessed during its trip through a common carrier's digital infrastructure.

Tariffed Telecommunications Services Magic Numbers

DS-X/TX Hierarchy

Existing longhaul common carrier circuits operate at throughput bit rates which are different from but compatible with uncompressed single or time division multiplied television signal bit rates. However, the switching nodes which pass signals from circuit to circuit can be literal *bit bottlenecks*. Some codecs in use by common carriers have circuitry that can switch 45 Mb segments of a high-speed digital transmission system, but they are the exception. One common carrier offers a user controllable 45 Mb service which can be used for coast-to-coast transportation of television signals. These services and promised future broadband transmission services are described in the next section.

Means for digital transmission of analog voice (telephone traffic) signals were originally developed by the Bell System two decades ago, to multiplex groups of voice bandwidth (30–3000 Hz) signals onto unshielded twisted pairs in high circuit count trunking cables. These time division multiplexed (TDM) systems offered two advantages over FDM systems: substantially more channels could be carried on each pair of wires and transmitted voice quality was dramatically improved by elimination of IMD and noise reduction.

Level 0 in the hierarchy of transmission speeds is 64,000 bps. This rate accommodates one 56,000 bps digitized voice channel, which is created by sampling the voice signal at a 4000 Hz rate and digitizing the sample to 7 bit resolution. This level is synonymously referred to as both "T0" (T-zero) and "DS-0," referring to publications of tariffs and digital signaling rates.

Level 1, popularly known as the *Bell System T1 Carrier,* has a digital signaling rate (DS-1) of 1.544 Mbps. It has a capacity of 24 voice channels. The only other widely tariffed level in this hierarchy available to broadcasters is DS-3 (T3), with a signaling rate of 44.734 Mbps (familiarly called 45 Mb.) It has a capacity of 28 T1 channels, for a total of 672 voice channels.

DS-2 services (6.312 Mbps signaling rates) are not currently widely tariffed. However, this rate is compatible with bit rates created by compression of video signals in MPEG-2 encoders.

Interfaces to T1 and T3 transmission services can be provided within a broadcast plant by several competing types of common carriers. These include the original 22 Bell System telephone companies (telcos), now called Regional Bell Operating Companies (RBOC) owned by the seven Baby Bells (RHCs—Regional Holding Companies), and by the growing number of dozens of new providers including independent telephone companies, alternative services providers, cable companies, wireless cable companies, regional cable interconnects, electric utilities, rail transportation companies, state highway departments and startup and established longhaul broadband transmission services providers.

All of these organizations also run fiber cables under city streets and into buildings. The wireless terrestrial and satellite microwave transmission services serve broadcasters via dishes on station towers and roofs.

Non-broadcasters utilize T0, sub-T1 (primarily 384 and 768 kbps), and T1 transmission services to establish point-to-point and point-to-multipoint videoconferencing networks at the lowest possible line costs. Broadcasters currently rate the received quality of the compressed bandwidth NTSC full motion video used for video conferencing between terrible and unacceptable for the purpose of broadcast signal distribution. These services can nevertheless be most acceptable in the television production and broadcasting industry for the slow scan transmission of still images.

Further, sub-T1 may one day become acceptable for distribution of MPEG-2 compressed network and syndicated programming to TV stations and cable head ends delivering several compressed standard definition (SD) signals over their 6 MHz channels.

If T3 transmission of NTSC video/stereo audio television signals is to become acceptable for network program distribution as well event backhaul, a T3 television recording standard is also needed. One encoding/decoding cycle of T3 signal may now be deemed acceptable. But second or third or more codec iterations will inevitably occur when T3 backhauled programs are retransmitted on a T3 distribution network, or re-recorded down the network to accommodate local scheduling needs. Tandem multiple codecs create their own objectionable artifacts. They could be avoided by establishing hamstringing T3 re-recording practices or reduced by recording all T3 backhauled programming onto a T3 digital television recorder for editing it in a T3 digital video/audio editing suite—which presently no manufacturers in the broadcast industry have shown any interest in producing.

Broadband Integrated Services Digital Network

Integrated services digital network (ISDN) is a narrowband service. Its capabilities that are not related to broadcasters' full motion, color video/surround sound audio transmission needs in any easily understandable way. Its three channels require a digital terminal, originally introduced in the $1000 price range, for their creation. Described as "*2B + D*," the "B" channels operate at 64 kbps and the "D" channel at 16 kbps. The B channel traffic can be either digitized voice, a high-speed digital transmission or the slow-scan transmission of a single frame of video.

Grouping of 2, 4 or 6 B channels yields transmission rates of 128, 256 and 384 kbps, providing faster transmission of higher resolution stills and higher frame rate motion image sequences. They should find favor for use in remote archive browsing, if the circuits can be user-established, dialed up and bandwidth expanded on demand, and are as available as DDD telephone service.

Switched Megabit Data Service (SMDS)

This service's name once invoked the inference that it would have applications for broadcasters in at least local and regional dialup switched transmission of television signals.

Synchronous Optical Network (SONET)

Hardware already exists for converting a 45 Mbps T3 signal into a synchronous transport signal level 1 (STS-1) electrical signal for transportation on an OC-1 carrier. Higher levels of the hierarchy of interest to broadcasters and HDTV advocates are OC-3 at 155.52 Mbps (accommodating a full bandwidth NTSC video/4-channel audio signal D2 and proposed D3 140 Mbps signal), and OC-12 at 622.08 Mbps (accommodating digitized wide-screen/high resolution images and accompanying sound with little or no compression). Some of the half dozen companies now advertising the availability of SONET hardware even predict the near future availability of bandwidth on demand services.

As this book went to press, this truly broadband (by television industry definition) service is still more of a future promise than a reachable reality. A new service at 270 Mbps has been offered in some areas which may offer a better cost-performance trade-off for broadcasters.

6.11 SATELLITE EARTH STATIONS AND SYSTEMS

JAMES H. COOK, JR.
SCIENTIFIC-ATLANTA, INC., ATLANTA, GA

INTRODUCTION

The use of satellite technology for communications is highly developed and in widespread use throughout the world. In three decades, it has progressed from a technology of limited acceptance to one of routine provision of services, including television and radio broadcast services. This acceptance has been based on the unique characteristics of satellites placed on the geosynchronous satellite orbit. The capability of a single quasi-stationary repeater in the sky, visible to large, contiguous regions, offers unique distribution capabilities for broadcast services. Satellite communications is also particularly useful for long distance communication services, for services across oceans or difficult terrain and for point-to-multipoint services.

Satellites in the geosynchronous orbit rotate from west to east. They appear fixed in space to earth stations on the ground because they orbit in synchronism with the earth's rotation. A satellite that is closer to the earth orbits faster; one that is beyond synchronous orbit rotates slower than the earth. Compare the 90 minute orbit of the Space Shuttle, that operates roughly 150 miles above the earth, with the 28 day orbit of the moon. Satellites located in the 22,300 mile high geosynchronous orbit have direct lines of sight to almost half the earth, as shown in Figure 6.11-1, therefore geostationary satellites are in effect unmanned relay stations. Except for small regions near the North and South Poles, widely separated earth stations can be seen from a single satellite.

Communications by satellite was made possible by parallel advances in space technology and electronics. In 1945 Arthur C. Clarke, the noted British scientist and science fiction writer, proposed relay stations in geostationary orbit for satellite communications. It took until 1963 for advances in technology, in solid-state electronics and in the thrust capability of rockets, to allow the placing of a satellite into a stationary orbit.

Communication by satellite is completely different from that by long-distance radio. Long distance communication at radio frequencies is possible because the *ionosphere*, produced by bombardment of the upper atmosphere by the sun, usually acts as a mirror to

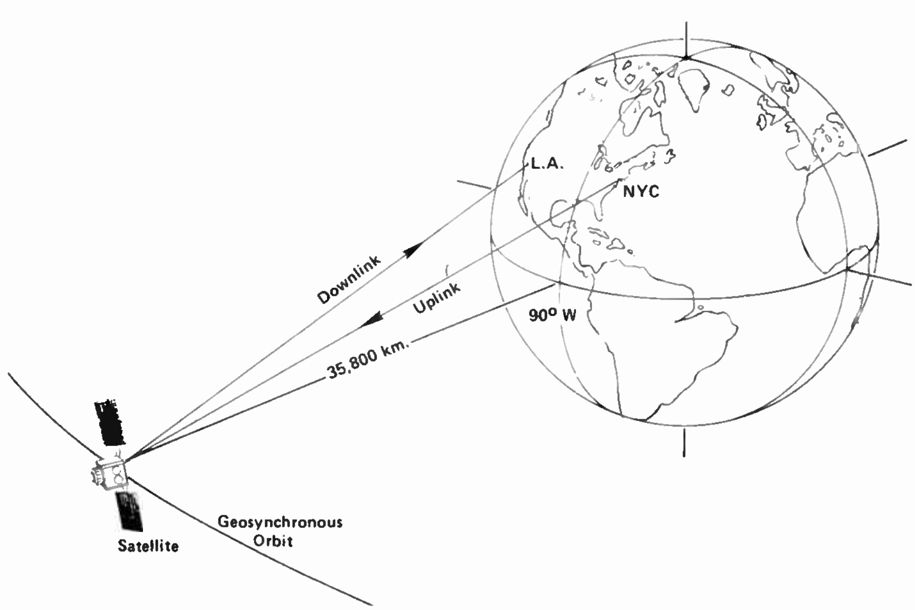


Figure 6.11-1. Satellite in geosynchronous orbit.

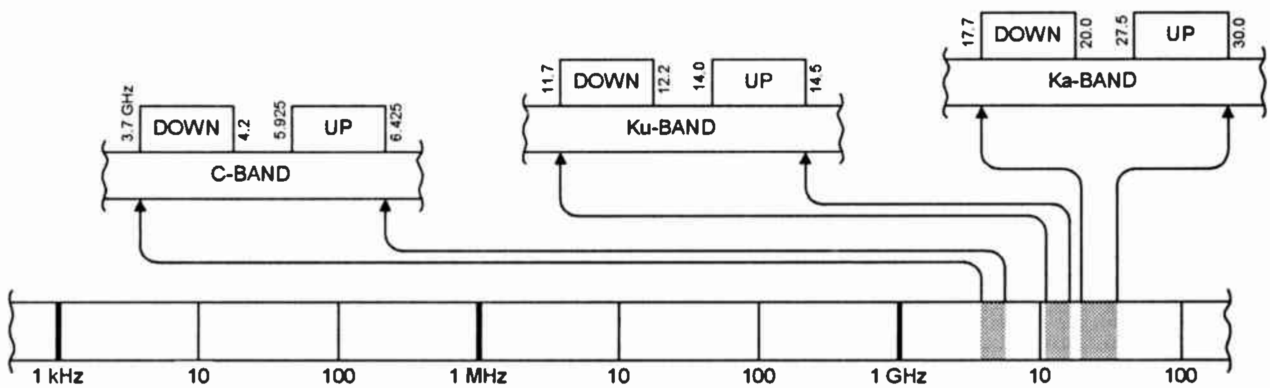


Figure 6.11-2. Frequencies in the microwave range of communications satellites.

reflect certain radio waves back to earth. As the frequency increases, a critical point is reached where the ionosphere ceases to act as a reflector, letting the waves pass through into space. Of course, television signals in the frequency range above 54 MHz do not usually lend themselves to long distance transmission. Therefore, long distance transmission of television signals was accomplished by either coaxial cable or terrestrial microwave links prior to the advent of satellite communications. Transmission of radio frequency, (RF) signals through the medium of fiber optics offers another choice of television distribution today.

Frequency

Communication satellites operate at microwave frequencies as shown in Figure 6.11-2. At microwave frequencies the ionosphere is always virtually transparent regardless of sunspot activity or time of day, permitting continuous, almost loss free transmission to and from satellites in orbit. In the United States the domestic commercial communications satellite networks operate in the fixed satellite services (FSS) frequency bands as defined by the Federal Communications Commission. Most of the domestic systems operate in either the C-band (6 and 4 GHz) or Ku-band (14 and 12 GHz) frequency ranges with C-band generally preferred because of superior propagation characteristics. The Ka-band frequencies (30 and 20 GHz) are also set aside for FSS operation and the use of these frequencies is expected to become quite prevalent in the early 2000s. The primary use in the United States to this date has been with experimental satellites (NASA Advanced Communications Technology Satellite, ACTS).

International systems provide services on a global basis (global beams) to all countries visible from a single orbit location and on a regional basis with spot beams. *Intelsat* and Intersputnik are examples of this type of system, using both C-band and Ku-band. The international satellite communication frequency bands are similar to the U.S. frequencies at C-band, but somewhat different at Ku-band. The frequency bands are determined by joint negotiations by the countries of the world through the auspices of the Radio Communications Sector of the International Telecommunications Union (ITU-R). International systems have begun to use an extended C-band frequency range of 3.4 GHz to 4.2 GHz rather than the 3.7 GHz to 4.2 GHz US band as well as 11.2 GHz to 11.7 GHz for downlink transmissions.

Satellite Stationkeeping

It was stated that the synchronous satellite appears stationary in space. Actually, a synchronous satellite is never perfectly stationary, because a number of forces including the pull of the sun and the moon perturb its orbit. If left alone, the satellite would eventually drift out of orbit. To overcome this, the position of the satellite is continuously monitored by an earth station, called a telemetry, tracking and command (TT&C) station and small jets of propellant such as hydrazine are used to keep it in position within a *station-keeping box*. The station-keeping box is typically a square $\pm 0.1°$ or less on each side and oriented with the sides parallel and perpendicular to the orbital plane. Sufficient rocket propellant must be carried on board to last for the satellites, predicted life, usually from 7 to 10 years. In the international arena, *COMSAT* has introduced a technique to extend the life of a satellite during its later years.[1] This technique, the *COMSAT Maneuver,* allows the satellite to drift north and south an increasing amount as the satellite approaches its end of life with excursions of several degrees by the time the rocket propellant is entirely spent. The *COMSAT Maneuver* increases the service life of the satellite by a number of years, but places the burden of tracking the satellite position on the earth station. Recently the *COMSAT Maneuver* has been used with some U.S.

[1] COMSAT is a publicly held satellite communications company, created by an act of Congress in 1963, and historically has represented the U.S. in the international satellite consortia Intelsat and Immarsat.

satellites at or near their end-of-life operation, resulting in significant economic benefits.

Satellite Footprint

The transmitting and receiving antennas on the satellite are designed to cover only specific desired regions of the earth's surface. This has several purposes. It concentrates the power radiated from the satellite into the desired direction, increases the sensitivity of its receiving antennas and helps prevent interference with signals from other satellites. The part of the earth's surface covered by a satellite is called the satellite's footprint. The footprint may cover one or more relatively localized regions of the earth or almost a complete hemisphere. A typical footprint is shown in Figure 6.11-3. The footprint is, of course, not sharply defined. Signal strengths tend to peak near the center of the footprint and roll off steeply past the 3 dB contour. Global or regional beams are usually shaped such that a particular defined section of the earth's surface is illuminated by the satellite's radiated signal. For example, a United States domestic satellite would probably limit the footprint such that its neighboring countries were illuminated with minimum signals to minimize interference possibilities. This is necessary since several countries share the same portions of the orbital arc.

Polarization

Electromagnetic waves and antennas are always *polarized* in some manner. The polarization may be linear, circular or elliptical. For the purposes of discussion elliptical polarizations will be dismissed as being nonideal cases that are intended to be either linear or circular. Linear polarizations and circular polarizations are aligned in space as shown in Figure 6.11-4. A linearly polarized antenna receives maximum power from an incident linearly polarized wave if the *tilt angles* of the wave and the antenna polarizations are aligned in space as in Figure 6.11-4(a). The wave is then said to be *co-polarized* or *polarization matched.* As the tilt angle of the wave or antenna rotates from co-polarization, the received power decreases. When the tilt angles are 90° apart as shown in Figure 6.11-4(b), the antenna is *cross polarized* to the wave and receives no power from it. The antenna and the wave then have *orthogonal* polarizations. A given wave can have two orthogonal polarizations which exist simultaneously and carry different information without interference. This principle, *frequency reuse,* is used to increase the "information capacity" of satellites and of the geosynchronous orbit.

Circular polarizations have either right hand (RHC) or left hand (LHC) senses. RHC and LHC polarizations are orthogonal. A circularly polarized satellite and a circularly polarized earth station are co-polarized if they have the same senses and are cross-polarized if they have the opposite senses. The relative tilt angles of circular polarized antennas and waves are of no consequence and are not defined. This represents an advantage of circular polarization over linear polarization since the tilt angle of the earth station does have to be adjusted for a particular satellite. On the other hand, circularly polarized antennas tend to cost more than linearly polarized antennas due to the increased complexity of the feed components. Most domestic satellites are linearly polarized while INTELSAT satellites are for the most part circularly polarized.

SATELLITE SYSTEM CHARACTERISTICS

The design of a satellite communication system is an intricate process, involving trade offs between many variables to obtain maximum performance at a reasonable cost. The major cost and complexity trade off occurs between satellite and earth stations, or more

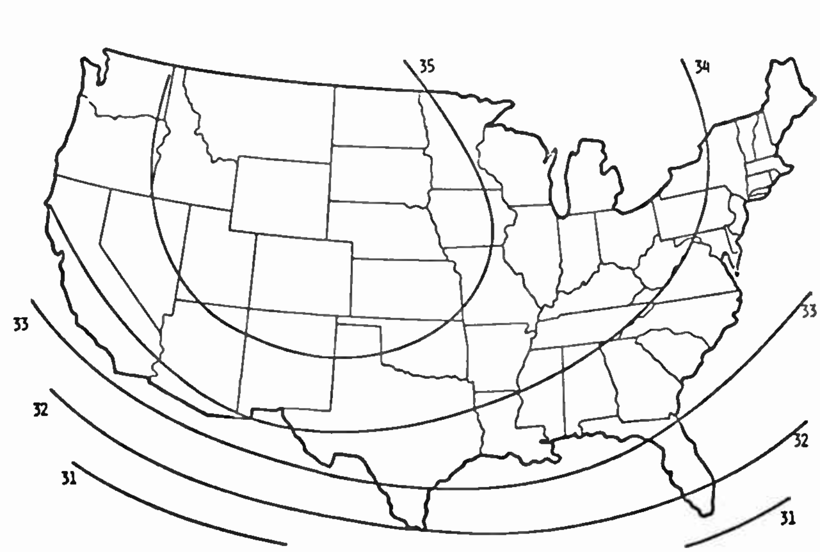


Figure 6.11-3. Satellite footprint.

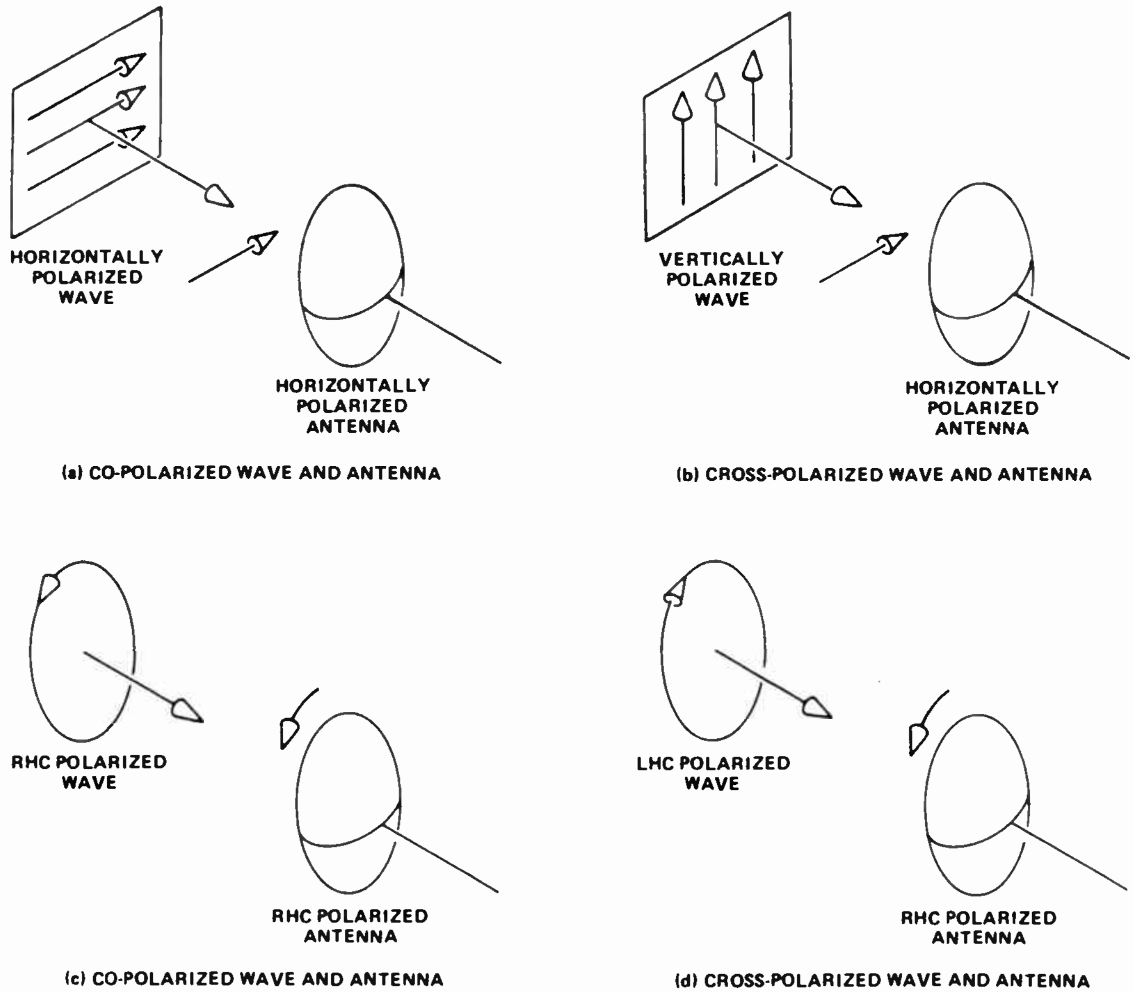


Figure 6.11-4. Linear and circular polarizations; a) co-polarized wave and antenna; b) cross-polarized wave and antenna; c) co-polarized wave and antenna; d) cross-polarized wave and antenna.

generically, between space segment and ground segment. The dominating design factors in both segments for systems using geostationary satellites are:

Space Segment

- Weight and size of satellite
- DC power generated on board
- Dimensions and complexity of satellite antennas

Ground Segment

- Allocated frequency bands
- Earth stations antenna size and RF capabilities
- Earth stations multiple access techniques

The weight of the satellite is limited by the high cost of launching a spacecraft into geostationary orbit, typically $40,000–$50,000 per kilogram. For a satellite of limited weight and size, a limited number of solar cells can be deployed which defines an upper limit on the dc power available for the communication transponders. The size and power limitations translate into the fact that the spacecraft has a limited RF output power, which then must be transmitted onto particular areas of the earth, (e.g., the continental United States).

Furthermore, power densities over the earth's surface are limited, depending on operating frequency bands, to allow interference free coexistence with terrestrial systems operating in the same frequencies. The result of all of this is the signals arriving from communication satellites are inherently weak, typically -120 to -160 dBW/m^2 and therefore relatively large receiving ground antennas must be utilized.

Multiple access and multiple destinations are distinctive virtues of satellite communications. The methods by which a large number of earth stations share one satellite or one transponder providing the required connectivity (multiple access techniques), also have a significant impact on system design. The multiple access can be achieved by sharing the transponder bandwidth in separate *frequency* slots, frequency division multiple access (FDMA), or the transponder availability in discrete *time* slots, time division multiple access (TDMA). A third technique, code division multiple access (CDMA) or spread spectrum, shares the transponders by allowing *coded* signals to overlap in time and frequency.

A satellite communication system must be designed to meet certain minimum performance standards, within limitations of transmitted power, RF bandwidth and antenna sizes. The most important performance criterion for analog systems is the signal-to-noise ratio (SNR or S/N) in the information channel or baseband. In digital systems the performance measurement criterion is bit error rate (BER).

SNR and BER depend on a number of factors, such as the predetection (incident upon the receive antenna) carrier-to-noise-density ratio (C/N_0) and the carrier-to-noise ratio (C/N) in the receiver, the type of modulation and the RF and baseband bandwidths. In the following section the design and analysis of satellite communications links in terms of C/N_0 will be conducted. Therefore, the carrier power received in an earth station receiver and the noise power density in the receiver need to be calculated to establish the operating link C/N_0.

Since satellites are inherently power limited, there will be invariably a modulation technique whereby a trade off of bandwidth for power will make the baseband S/N larger than the RF C/N in analog systems and will optimize BER in digital systems.

Satellite Transmission Modes for Television

Analog and digital formats are both used for the transmission of television signals via satellite. FM has been the prevalent analog technique because:

- It minimizes the effects of non-linearities in the transmission channel
- It is immune to AM noise
- Power-limited systems can take advantage of the wider bandwidth to increase the C/N
- Various processing techniques can be employed to optimize video transmission (multiplexing, preemphasis and threshold extension)

The choice of the optimum modulation index **m**, the ratio of the FM deviation to the highest modulating frequency f_m, is critical. Bandwidth and deviation are related by Bandwidth = $2(\mathbf{m} + 1)f_m$. The spectral distribution of a FM signal as a function of **m** is shown in Figure 6.11-5. The selection of the optimum deviation must be based on the number of channels to be transmitted, the type of baseband signal (component or composite), the signal quality requirement, the power received and the available bandwidth. A typical value

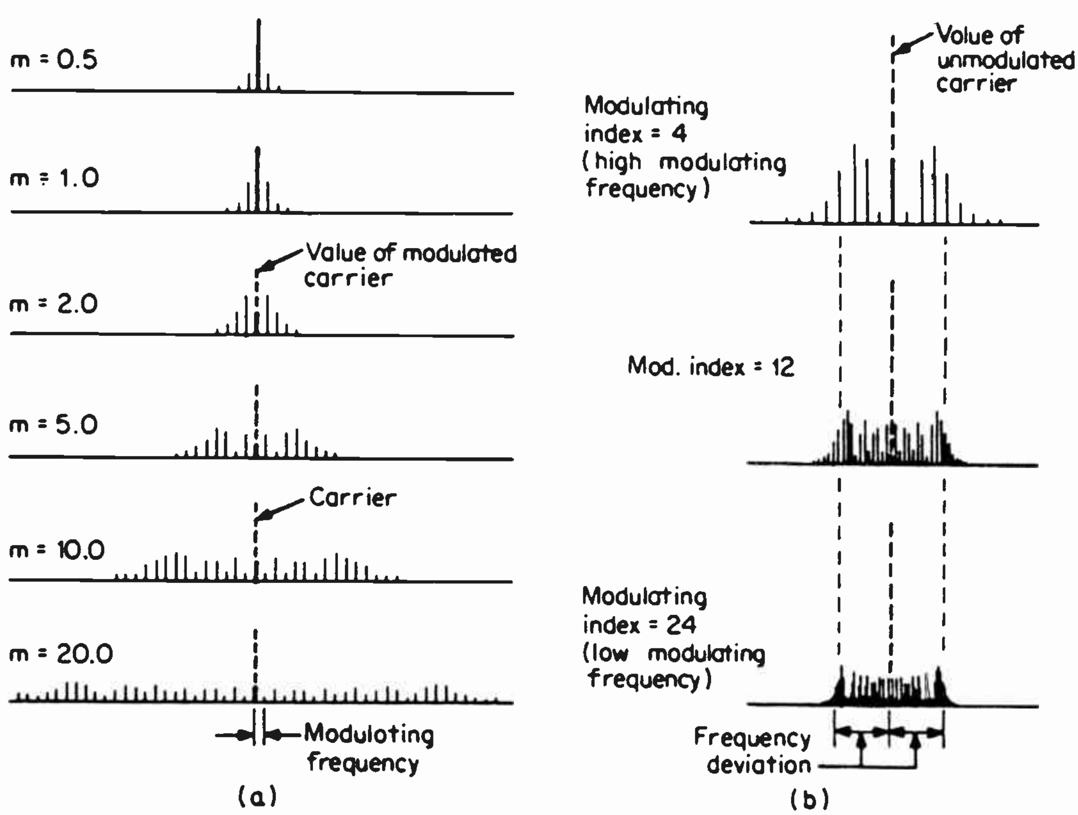


Figure 6.11-5. Spectra of frequency modulated signals a) frequency spectra with increasing frequency deviation and constant modulating frequency; b) frequency spectra with constant frequency deviation.

of peak Δf for a C-band satellite transponder with a nominal bandwidth of 36 MHz is 10.75 MHz.

For satellite transmission, in addition to the deviation of the carrier by the signal, it is usually necessary to subject the main carrier to a low-frequency deviation. This spreads the high concentration of carrier and sideband energy over a larger range of the spectrum and permits higher satellite effective radiated power (ERP) without exceeding the FCC's limit on watts/meter^2/kHz downlink power density.

Preemphasis/deemphasis is employed in FM systems for the transmission of video to compensate for the increase in thermal noise with increasing frequency. CCIR Recommendation 567 specifies a standard 75 μsec preemphasis, for example.

A characteristic of FM is that the detected S/N for the video signal is higher than its C/N ratio. This difference is the FM improvement factor; satellite transmission takes advantage of this improvement, provided the received C/N is greater than the receiver operating threshold.

Threshold extension demodulation (TED) is a common technique used in FM receivers to reduce video impulse noise when the C/N drops below the receivers's operating threshold. Above threshold, the receiver acts like a standard discriminator; when C/N drops below threshold, TED circuitry automatically switches to a narrow bandwidth.

Digital Transmission

Satellite systems are used for the transmission of digital information. A typical digital transmission uplink consists of a modem, upconverter and power amplifier. The modem converts digital information to and from a modulated carrier. The center frequency of the modulated carrier positions the signal within a satellite transponder. The upconverter converts the modulated-carrier to a satellite frequency and thus selects the transponder of the satellite.

The earth station component used to convert digital information to a format suitable for transmission by satellite is referred to as a modem (an acronym for modulator/demodulator. The modem accepts a digital data input signal and outputs an intermediate frequency, typically a range centered on either 70- or 140-MHz, containing the modulated digital information.

A modem for use in satellite systems is similar to modems used for telephone circuits. However, satellite modems generally operate at much higher bit rates and, in addition, contain special features specifically for satellite link use.

Transmitted digital data is first applied to the encoder section of the modem for forward error correction (FEC) encoding. This process adds redundancy to the bitstream, by appending additional bits to the original information, to provide error detection and correction capability. The signal is then scrambled using a standard algorithm to ensure random data. This ensures that the signal spectrum will be noise-like in nature, and minimize the possibility of spinious energy from transmission.

The aggregate data (original data plus error correction bits) is applied to the modulator for modulation onto an IF carrier. The IF carrier is selectable, typically in the range of 50 to 90 MHz (for 70 MHz operation), or 100 to 180 MHz (for 140 MHz operation). The center frequency of the modem modulator is tuned to position the signal within the satellite transponder.

A satellite digital transmission system is characterized by data interface, data rate, code rate and modulation scheme. Data rate refers to the number of bits per second transmitted by the modem. The data rate is typically front panel selectable. Typical ranges of rates are from 32 kbps to 3 Mbps selectable in increments as small as 1 bps. Modems typically support a number of data interfaces. The data interface refers to the connector and signal levels. Typical data interfaces are DS1, CEPT, EIA-422, V.35 and MIL-188/114. Code rate refers to the FEC-encoding scheme. In some modems, the code rate is selectable. The code rate configuration is referred to as m/n. "m" refers to the number of original bits per block of transmitted bits. "n" refers to the number of original bits plus error correction bits per block of transmitted bits. Thus, a code rate of 3/4 means that for every input three data bits, four data bits are transmitted. Thus a 1024 kbps modem operating with a code rate of 3/4 would transmit 1365 kbps over the satellite channel.

The modulation scheme refers to the method of indicating data bits. Two common modulation schemes employed in satellite transmission systems are binary phase shift key (BPSK) and quadrature phase shift key (QPSK). These modulation schemes generate periodic phase shifts referred to as *symbols.* The symbol rate (the number of symbols per second) and data rate determines the amount of bandwidth required in the channel.

In BPSK, two phase shifts are used to represent two unique states. For this case the symbol rate is equal to the transmission rate. The QPSK scheme uses four phase shifts, thus transmitting two bits per symbol. For QPSK, the symbol rate is equal to half the transmission rate. QPSK requires less bandwidth than BPSK, but requires increased performance from the channel.

Advantages of Digital Television

The advantages of digital technology for television processing, including production, storage and distribution will make its use dominant in coming years. These advantages are: perfect multigeneration reproducibility, precise time and level controls, digital storage and signal processing, data compression, easy manipulation by computers and generation of multimedia constant, incorporation in digital transport packets for packet switching (ATM), broadband digital communication (B-ISDN) and, error control.

The 1982 adoption of ITU-R Recommendation 601 (CCIR601), for component television and the adoption of MPEG-2 by the International Standards Organization (ISO) and the International Electrotechnic Committee (IEC) in 1993 were major milestones for the industry.

The use of digital compression techniques, removing statistical redundancy (entropy coding) and taking advantage of psychophysics to remove irrelevant information (perceptual coding), allows the data rate of high quality television to be reduced by significant factors (~50) making its use very attractive in conservation of bandwidth. This technique is being used to deliver high quality television to millions of subscribers via direct broadcast satellite today.

C-Band Satellites

C-band was initially favored for communications satellites because of the favorable propagation characteristics at these frequencies. The specific bands in most common use are the 5925 MHz to 6425 MHz (uplink) and the 3700 MHz to 4200 MHz (downlink) band pair. U.S. domestic FSS requires the use of 36 MHz bandwidth channels placed on 40 MHz centers. A satellite using a single polarization can provide 12 such transponders, although all new satellites are mandated to be frequency reuse and provide 24 such transponders. Frequency reuse is implemented by the use of orthogonal polarizations and by staggering the microwave carriers of alternate transponders. As an example of a typical satellite, the transmit and receive frequency plans of a GE/RCA Satcom satellite are shown in Figure 6.11-6. The numbered brackets represent each channel. The bandwidth of the channel is represented by the width of the bracket. The carrier frequency, shown above the channel number, is centered on each channel. The signals of alternate transponders in the frequency plan of Figure 6.11-6 are nominally orthogonal. If they were exactly orthogonal and the associated earth stations were ideal (with respect to polarization), there would be no interference caused by the overlapping sideband energy of adjacent transponders. In practice, the polarizations of the antennas of the satellite and earth stations are not ideal. Some small amount of interference occurs, but the combination of nearly orthogonal polarizations and the use of the staggered frequency plan provides for high quality transmission under almost all weather conditions.

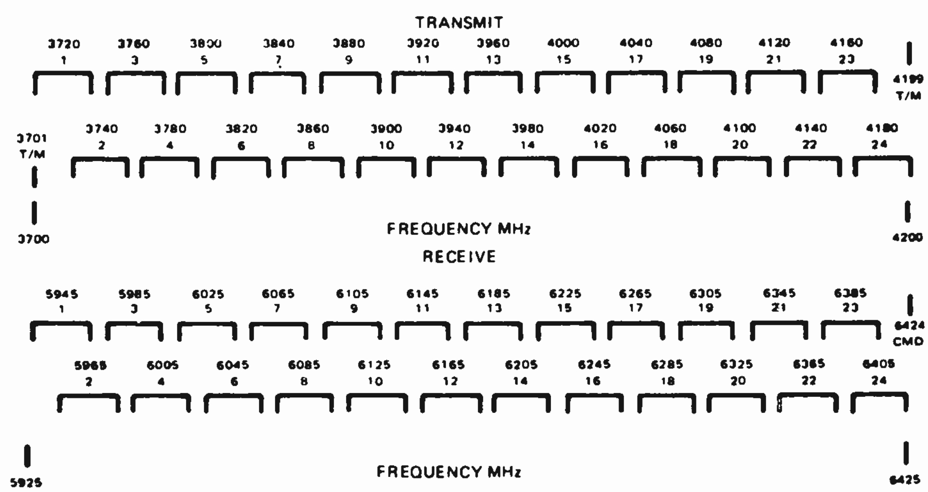
Ku-Band Satellites

The first systems using the 14.0 GHz to 14.5 GHz (uplink) band and the 11.7 GHz to 12.2 GHz (downlink) band were launched in 1976 by Satellite Business Systems (SBS). The higher propagation loss characteristics at these frequencies require higher spacecraft equivalent isotropic radiated power (EIRP) to achieve the same transmission performance as C-band frequencies and this is obtained from the use of greater spacecraft antenna gains, readily achievable at the higher frequencies. Since the Ku-band frequencies are not shared with terrestrial systems (as is the case of C-band) the power flux density (PFD) limitation is less stringent and there is no requirement for coordination with terrestrial microwave systems. The high powers permit the use of very small earth station antennas at or near the user's premises. This results in important economic advantage for many services and makes the use of this frequency band very attractive. Even so, a good part of the higher satellite power achievable is necessary to offset the additional attenuation that is experienced at these frequencies during heavy rain conditions.

There is no mandated frequency plan for transponders in this frequency band, although typical transponder bandwidths are 36 to 72 MHz. Since the bandwidth is the same as C-band it is possible to have a similar 24 transponder, 36 MHz frequency plan with 40 MHz channel spacings when frequency reuse is utilized.

Two of the more important differences between C-band and Ku-band are the following:

- C-band FSS share frequencies with terrestrial microwave systems. This places constraints on the location



NOTE: NUMBER BELOW CHANNEL CENTER FREQUENCY REFERS TO TRANSPONDER IDENTITY.

Figure 6.11-6. GE/RCA Satcom satellite frequency plans.

of C-band earth stations and it limits the permissible downlink power density.

- Ku-band signals are subject to significant attenuation in heavy rainfall

The advantages and disadvantages of C-band and Ku-band, which result from these and other differences, are summarized in Table 6.11-1.

Ka-Band Satellites

The commercial application of Ka-band satellites is only in its infancy years. The NASA ACTS experimental satellite has allowed successful demonstration of the use of the 17.5 GHz to 22.5 GHz downlink and 29.5 GHZ to 34.5 GHz bands for various services. It is inevitable that these bands will be used in the future for video applications, with considerable interest in narrow casting for particular regions of the country.

In mid-1998, the FCC granted licenses to thirteen companies for 73 geosynchronous satellites which will operate in the Ka-band.

Table 6.11-1
Merits of C- and Ku-Band for Satellite Communications.

C-band Advantages
Less susceptible to rain outages. Established manufacturing infrastructure. Antenna surface tolerance can be achieved by various techniques that lend themselves to low cost manufacturing.
C-band Disadvantages
Frequency band is congested because it is shared with terrestrial microwave, making frequency coordination a requirement. Requires relatively large antennas because of low satellite EIRP levels and the necessity of narrow half-power beamwidth to allow two degree spaced satellites. Avoiding terrestrial interference can make site selection a difficult process. The use of artificial shielding to block interference can increase total system cost. Faraday rotation of polarization can affect system performance. Satellite dispersal signal is required to prevent harmful interference to terrestrial stations, resulting in more stringent video receiver clamping specifications.
Ku-band Advantages
Frequency band is only used for satellite communication. Smaller antennas may be used because of higher gain and higher satellite EIRP. Easier site selection because of smaller size of antenna and lack of terrestrial interference. Narrower antenna beamwidth is desirable in reduced orbital spacing Suitable for direct-to-home application. Lower reception equipment cost. Flexibility in channelization plan. Not affected by Faraday rotation. No satellite dispersal signal disadvantages.
Ku-band Disadvantages
Affected by rain attenuation and depolarization. Narrow beamwidths of antennas may require more rigid mounts Reflector surface tolerance requirements restrict manufacturing techniques and increase cost. Waveguide and coaxial transmission line losses are quite high. Noise temperature of Low Noise Amplifiers may cause the use of large antennas to achieve desired G/T.

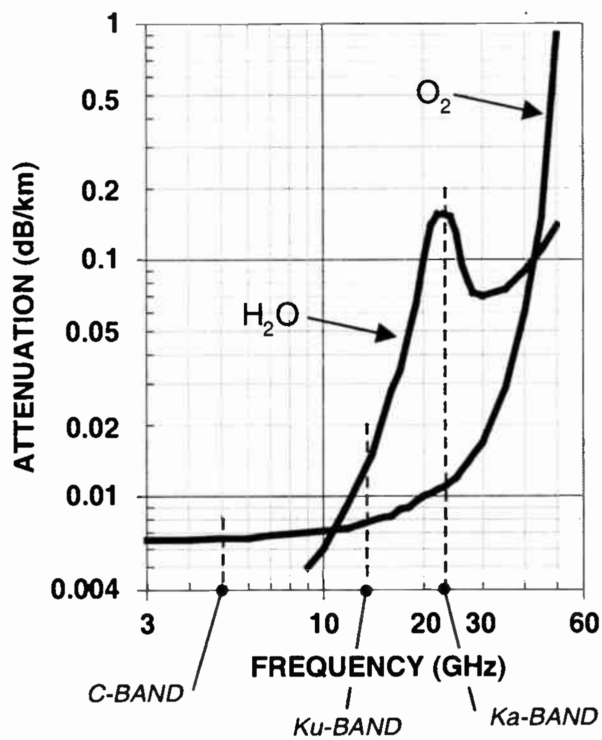


Figure 6.11-7. Signal attenuation versus frequency due to atmospheric oxygen and rainfall.

The Ka-band, in the context of fixed-and mobile-satellite communications services, refers to the downlink frequency band of 17.7–20.0 GHz, and the uplink frequency band of 27.5–30.0 GHz. It is a new band in a regulatory sense, having been created in July 1996 when the FCC allocated these frequencies for this service. In early 1998, the satellite industry completed an agreement on an orbital location assignment plan for the Ka-band satellites being proposed, which was adopted by the FCC and followed by the issuance of licenses for satellite system construction and operation.

Ka-band signals suffer from greater attenuation due to the presence of rain and atmospheric oxygen, than do C-band and Ku-band services. This is illustrated in Figure 6.11-7, where the approximate center of each band is marked by a dashed line. This attenuation problem has historically made Ka-band rather unattractive for satellite communications, but scarcity of spectrum in other bands as well as advances in satellite communications technology have together brought about a change in that situation.

In addition to core satellite voice, data, and video offerings, Ka-band licensees will provide many low cost, global broadband interactive services such as:

- Desktop-to-desktop videoconferencing
- Direct-to-home video
- Distance learning and corporate training
- Internet access

- Telemedicine
- Electronic transaction processing
- Satellite news gathering

Regulatory Issues

Satellite communication systems are governed by the FCC in the United States and by the ITU on the international level. The governing agencies assign frequency bands of operation, satellite performance characteristics and orbit location and provide technical specifications of radiated power density and radiation gain patterns for the earth stations. The FCC is the licensing body for all transmit earth stations in the United States and licenses C-band receive only earth stations at the owner's request. The FCC Rules and Regulations, Part 25, form the basis of the applicable documents which must be followed for the planning and implementation of any FSS band satellite communication system.

The FCC amends and interprets the rules as the technology and the requirements of satellites changes through amendments, decisions and declaratory orders; therefore, it is recommended that the FCC be contacted at the time of system planning to obtain the latest rules and regulations. The FCC established precedents for the minimum diameter apertures and sidelobe gain envelopes for earth station antennas operating in the FSS bands at the beginning of these services in the early 1970s to minimize interference between terrestrial systems and satellite systems and between satellite systems. These precedents have been modified through the years as the use of satellite services has increased. The more significant recent rulings pertaining to earth station antenna performance have resulted in improved antenna radiation patterns in the close-in sidelobe region and have established maximum radiated power densities for antennas diameters less than 9 m in diameter for C-band operation and 5 m in diameter for Ku-band operation.

The FCC Rules and Regulations Part 25.209 pertaining to antenna gain envelopes is mandatory for all transmit antennas. Excerpts from this standard follow (refer to the current rules publication for the entire text):

a) The gain of any antenna to be employed in transmission from an earth station in the fixed satellite service shall lie below the envelope defined below:
   1. In the plane of the geostationary satellite orbit as it appears at the particular earth station location:

$[29 - 25\log(\theta)]$ dBi	$1° \leq \theta \leq 7°$
$+8$ dBi	$7° < \theta \leq 9.2°$
$[32 - 25\log(\theta)]$ dBi	$9.2° < \theta \leq 48°$
-10 dBi	$48° < \theta \leq 180°$

   where θ is the angle in degrees from the axis of the main lobe and dBi refers to the dB relative to an isotropic radiator. For the purposes of this section, the peak gain of an individual sidelobe may not exceed the envelope defined above for θ between 1° and 7°. For θ greater than 7°, the envelope may be exceeded by 10% of the sidelobes, but no individual sidelobe may exceed the envelope by more than 3 dB.
   2. In all other directions:
   Outside the main beam, the gain of the antenna shall lie below the envelope defined by:

$[32 - 25\log(\theta)]$ dBi	$1° \leq \theta \leq 48°$
-10 dBi	$48° < \theta \leq 180°$

   where θ is the angle in degrees from the axis of the main beam and dBi refers to dB relative to an isotropic radiator. For the purpose of this section, the peak gain of an individual sidelobe may be reduced by averaging its peak level with the peaks of the nearest sidelobes on either side, or with the peaks of the two nearest sidelobes on either side, provide that the level of no individual sidelobe exceeds the gain envelope given above by more than 6 dB.

b) The off-axis cross-polarization isolation of any antenna to be employed in transmission at frequencies between 5925 and 6425 MHz from an earth station to a space station in the domestic fixed-satellite service shall be defined by:

$[19 - 25\log(\theta)]$ dBi	$1.8° < \theta \leq 7°$
-2 dBi	$7° < \theta \leq 9.2°$

c) Any antenna licensed for reception of radio transmission from a space station in the fixed-satellite service shall be protected from radio interference caused by other space stations only to the degree to which harmful interference would not be expected to be caused to an earth station employing an antenna conforming to the standards defined in paragraphs a. and b. of this section.

d) The standards specified in paragraphs a. and b. of this section shall apply to all new earth station antennas initially authorized after February 15, 1985 and to all earth stations antennas after March 11, 1994.

e) The operations of any earth station with an antenna not conforming to the standards of paragraph a. and b. of this section shall impose no limitations upon the operation, location and design of any terrestrial station, any other earth station, or any space station.

f) An earth station with an antenna not conforming to the standards of paragraphs (a) and (b) of this section will be routinely authorized after February 15, 1985 upon a finding by the Commission that unacceptable levels of interference will not be caused under conditions of uniform 2° orbital spacings.

g) The antenna performance standards of small antennas operating in the 12/14 GHz band with diameters as small as 1.2 meters starts at 1.25° instead of 1° as stipulated in paragraph (a) of this section.

The FCC further acknowledged that the envelope defined above is only a reference envelope in the receive band. Receiving antennas do not have to conform to this envelope to be eligible for licensing. Facilities with performance worse than the reference envelope must, of course, be prepared to accept potentially higher interference levels. The interference levels should be calculated based on typical measured radiation patterns, site location and for a desired satellite or satellites. This analysis may result in acceptable receive-only carrier-to-interference performance for antennas meeting $[32 - 25\log(\theta)]$ envelope even with orbital spacings as small as 2°, since discrimination, that is, peak on-axis gain to sidelobe gain, is the important determining factor, not an arbitrary sidelobe gain performance envelope relative to isotropic.

FCC License

The FCC requires licensing of transmitting earth stations and permits licensing of receive only (RO) earth stations. It is desirable for a broadcaster to license a C-band RO earth station, since licensing protects the station from future interference from domestic microwave systems.

The FCC Rules and Regulations are ever evolving therefore it is strongly suggested that the FCC be contacted before filing to obtain the latest regulations and regulatory procedures effective at the time of the filing. For example, in December 1996, International Bureau Docket No. 95-117 *Report and Order* streamlined the Rules and Regulations for Satellite Application and Licensing Procedures. This Docket eliminated a series of previous forms that contained redundant information and simplified them into a single form, Form 312.

The application for a C-band RO terminal is filed on Form 312, Main Form and Schedule B. The earth station must conform to the provisions of Paragraph 25.131of the FCC Rules and Regulations. The filing requirements for a transmit station use the same Form 312, Main Form and Schedule B and must include the information specified in Paragraph 25.130 of the FCC Rules and Regulations.

SYSTEM PERFORMANCE ANALYSIS TECHNIQUES

Considering an RF link as illustrated in Figure 6.11-8 with transmit power P_t and transmit gain G_t, the effective isotropic radiated power, EIRP, for the station along the main beam of the antenna is the product $G_t{*}P_t$. At a distance R meters from the transmitter, the radiated flux density, S, becomes:

$$S = (G_tP_t)\frac{1}{4\pi R^2}k_a \qquad \text{Watts/m}^2 \qquad [1]$$

where k_a = atmospheric attenuation factor < 1.

If an antenna with an effective area in square meters, A_e, is receiving this flux density, the received carrier level, C, at the antenna output is:

$$C = SA_e = (P_tG_tA_e)\frac{1}{4\pi R^2}k_a \qquad \text{Watts} \qquad [2]$$

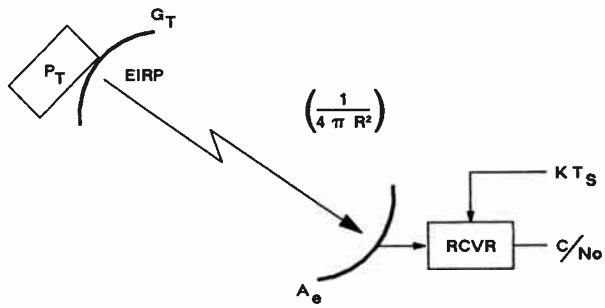


Figure 6.11-8. RF link diagram.

At the same antenna output point, the effective noise power density, N_o is given by:

$$N_o = kT_s \qquad \text{Watts/Hertz} \qquad [3]$$

where: k = Boltzmann's constant = $1.38{*}10^{-23}$ Joules/K or −228.6 dB.
T_s = System noise temperature in degrees Kelvin

We can assume the atmospheric attenuation factor to be unity with very small error (in particular for C-band and Ku-band links refer to Figure 6.11-7), therefore, the carrier-to-noise density ratio C/N_0 can be expressed by:

$$\frac{C}{N_o} = (P_tG_tA_e)\frac{1}{4\pi R^2}\frac{1}{kT_s} \qquad [4]$$

A fundamental relationship in antenna theory is that the gain, G_r and the effective area of an antenna, A_e, are related by:

$$A_e = G_r\frac{\lambda^2}{4\pi}\ \text{m}^2 \qquad [5]$$

Substituting this relation into the expression for C/N_o,

$$\frac{C}{N_o} = (P_tG_tG_r)(\lambda/4\pi R)^2\frac{1}{kT_s} \qquad [6]$$

or

$$\frac{C}{N_o} = EIRP\frac{G_r}{T_s}(\lambda/4\pi R)^2\frac{1}{k} \qquad [7]$$

The factor $(\lambda/4\pi R)^2$ is often inverted and defined as the spreading loss or space loss factor. This spreading loss can also be expressed as:

$$L_s = (4\pi Rf/c)^2 \qquad [8]$$

where: c = Speed of light = $3{*}10^{+8}$ meter/second
f = Frequency in Hertz

Link calculations are usually carried out in dB rather than directly from the above relations because of ease of working in common logarithms. C/N_o in dB can be calculated by:

$$(C/N_o)\ \text{dB} = 10\log(C/N_o)$$

$$(C/N_o)\ \text{dB} = EIRP - L_s + (G/T) + 228.6 \qquad [9]$$

where: $EIRP = 10\log(G_t P_t)$ dBW

$$L_s = 20\log(4\pi Rf/c)\ \text{dB} \qquad [10]$$
$$= 92.45 + 20\log R\ (\text{km}) + 20\log f\ (\text{GHz})$$

$$(G/T) = 10\log(G_r/T_s)\ \text{dB/K} \qquad [11]$$

Alternately, C/N_o can be expressed in terms of flux density, S, as:

$$(C/N_o) = S + (G/T) - A_i + 228.6\ \text{dBHz} \qquad [12]$$

$$S = EIRP - L_s + A_i\ \text{dBW/m}^2 \qquad [13]$$

where A_i is the effective aperture of an isotropic radiator in dB:

$$A_i = 10\log(4\pi/\lambda^2) \qquad [14]$$

Equation (9) is a fundamental tool for characterizing space link performance. It will be utilized later when calculating overall satellite link performance.

Earth Station Receive Figure of Merit G/T

G/T is the figure of merit of a receive system. It is primarily a function of the gain of the antenna along with the antenna noise temperature, first amplifier noise temperature and losses located between the antenna and the first amplifier. The importance of the term *G/T* in Equations [9] and [12] cannot be overstated. Examination of the C/N_o expression shows that for a given available transmitting power and information format (and thus bandwidth), the only available method of controlling the received signal quality that can be used by the downlink operator is through the system *G/T*. Note that the *G/T* provides a direct dB relationship with C/N_o.

Figure 6.11-9 shows a block diagram of a typical receive system. Each device in the RF path has an associated gain or loss and a noise temperature. These contributions are combined to reflect the noise power weighted by the gain distribution through the chain. The earth station *G/T* is given by:

$$(G/T) = G_a - 10\log(T_s)\ \text{dB/K} \qquad [15]$$

where: G_a = Antenna gain referenced to LNA input (dBi)
T_s = System noise temperature referenced to LNA input (K)

The system noise temperature, T_s, referenced to the LNA input can be calculated by adding as noise powers the equivalent noise temperatures of all noise contributors, weighted by the net gain between the point in which that noise is been added and the LNA input, that is:

$$T_s = (T_a/L_w) + T_o(L_w - 1)/L_w + T_{vswr} + T_{lna} + [(L_t - 1) + L_t(F_r - 1)]*T_1/G \qquad [16]$$

where: T_a = Antenna noise temperature (K)
L_w = Waveguide loss between antenna and LNA (linear power ratio)
L_t = Transmission loss between LNA and receiver (line power ratio)
T_o = Ambient temperature (K)
T_1 = 290 K
T_{lna} = LNA noise temperature (K)
T_{vswr} = LNA-Antenna impedance mismatch noise temp. (K)
F_r = Receiver Noise Figure (linear power ratio)
G = Net gain between LNA input and receiver input (linear, includes interconnect cable loss)

The antenna temperature is usually minimum at zenith, typically 15 to 25° for a low loss, C-band antenna with low wide angle sidelobes. As the elevation angle decreases, the antenna temperature increases because more of the higher level sidelobes look at the earth which has a temperature of about 290 K. A typical curve of the variation of noise temperature with elevation angle is illustrated in Figure 6.11-10. Similarly, Figures 6.11-11-a and 6.11-11-b show typical *G/T* system performance for different antenna diameters as a function of elevation angle for C-band and Ku-band, respectively.

Satellite Transponder

The orbiting spacecraft provides a one hop carrier relay over a wide geographic area. In C-band systems the uplink signal is transmitted near 6 GHz, received by

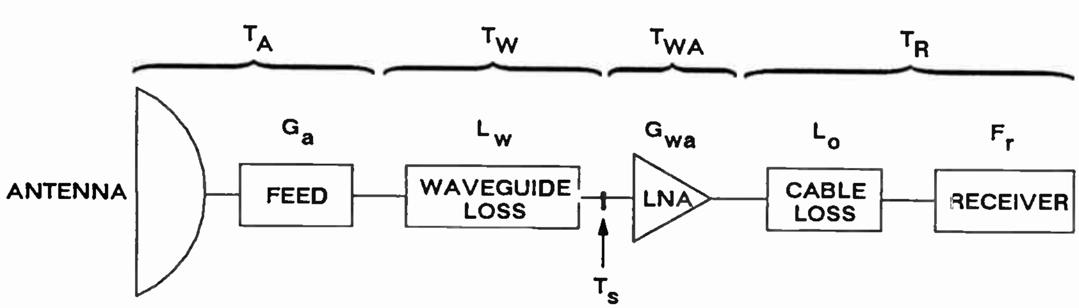


Figure 6.11-9. G/T system diagram: receive-only earth station.

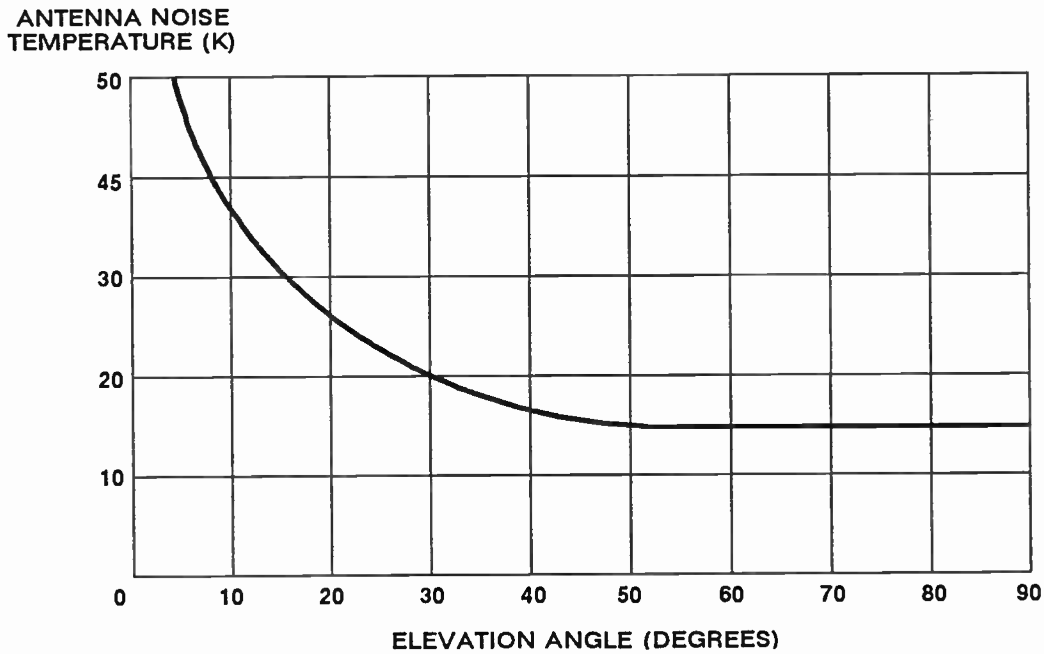


Figure 6.11-10. Typical antenna noise temperature variations with elevation angle.

the satellite, amplified, translated in frequency, filtered and retransmitted near 4 GHz. Likewise, in Ku-band systems the uplink occurs in the 14 GHz range and the downlink in the 12 GHz, and in Ka-band systems, the 30 GHz and 20 GHz ranges, respectively (see Figure 6.11-2).

Since the satellite serves as a transmit/receive station, it must be characterized by a *G/T* for the uplink side and by saturated EIRP for the downlink side. To couple the uplink and downlink signal strengths and as a definition of the transponder sensitivity, the uplink RF spectral flux density required at the satellite to saturate the transponder is also specified (saturation flux density or SFD). These three satellite parameters vary with geographic location. Contour maps called footprints are usually available for assessing these variations. Typical footprints for C-band and Ku-band satellites are shown in Figures 6.11-12 and 6.11-13, respectively.

Another important parameter which characterizes the transponder performance is the input/output power transfer and the intermodulation response. Both performance parameters are normally specified in terms of input back-off (BO_i) and output back-off (BO_o) back-off, that is, as a function of the power reduction expressed in dB with respect to saturation. Figures 6.11-14 and 6.11-15 show typical transponder response for a satellite equipped with a travelling wave tube (TWT) type power amplifier.

Satellite Link Analysis

With the preliminary procedures and formulations described previously, link calculations can be conducted. First, the distance or slant range from the satellite to the earth station needs to be determined so that the space loss may be calculated. From orbit geometry and Equation 10 above, the space loss expressed in dB is found to be:

$$L_s = 185.05 + 10\log[1 - 0.295\cos(H)\cos(\Delta L)] + 20\log f \qquad [17]$$

where: H = Latitude of earth station
ΔL = Difference in longitude for earth station and satellite
f = Frequency in GHz

The overall satellite link can now be calculated.

Uplink C/N

From equation (9), the uplink $(C/N_o)_u$ becomes:

$$(C/N_o)_u = EIRP_u - L_u + (G/T)_s + 228.6 \qquad [18]$$

or

$$(C/N_o)_u = S - A_i + (G/T)_s + 228.6 \qquad [19]$$

and

$$S = SFD - BO_i \qquad [20]$$

where: S = Flux density (dBW/m^2)
L_u = Uplink space loss (dB)
$EIRP_u$ = Uplink EIRP (dBW)
$(G/T)_s$ = Satellite G/T (dB/K)
A_i = 21.5 + 20 log f (GHz) (dB/m^2)
SFD = Saturation flux density (dBW/m^2)
BO_i = Transponder input Back-Off (dB)

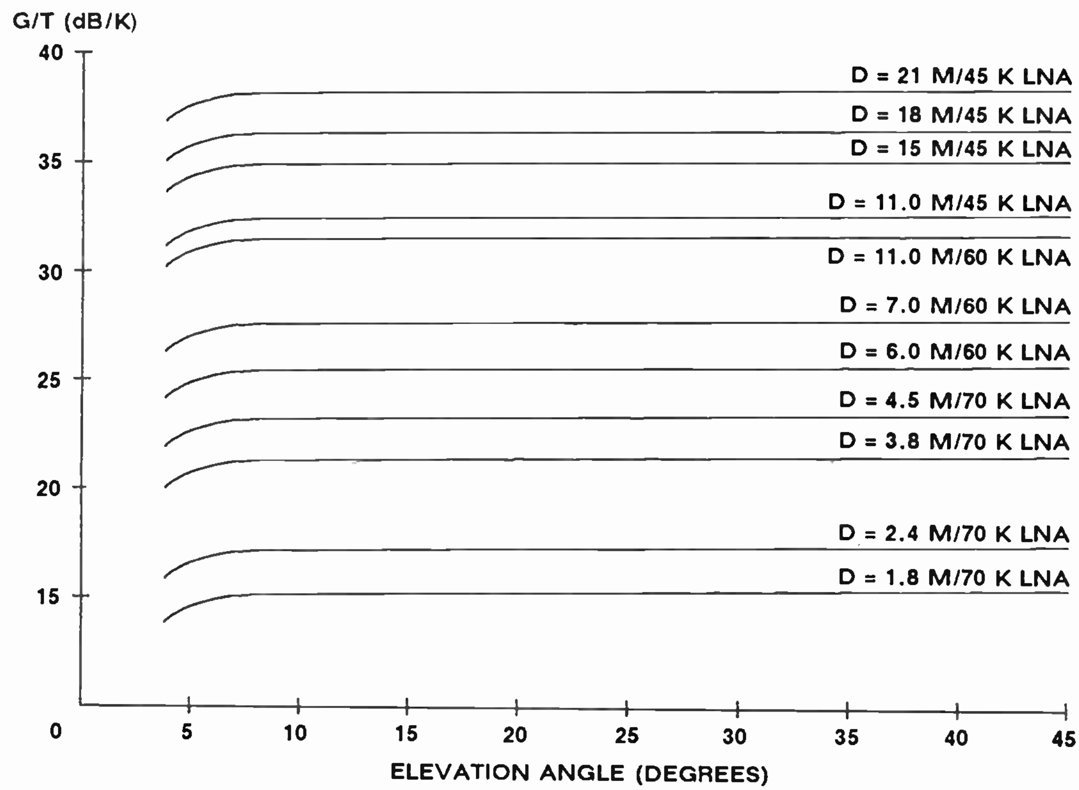


Figure 6.11-11a. Typical C-band g/T system performance versus elevation angle for different commonly used antenna diameters.

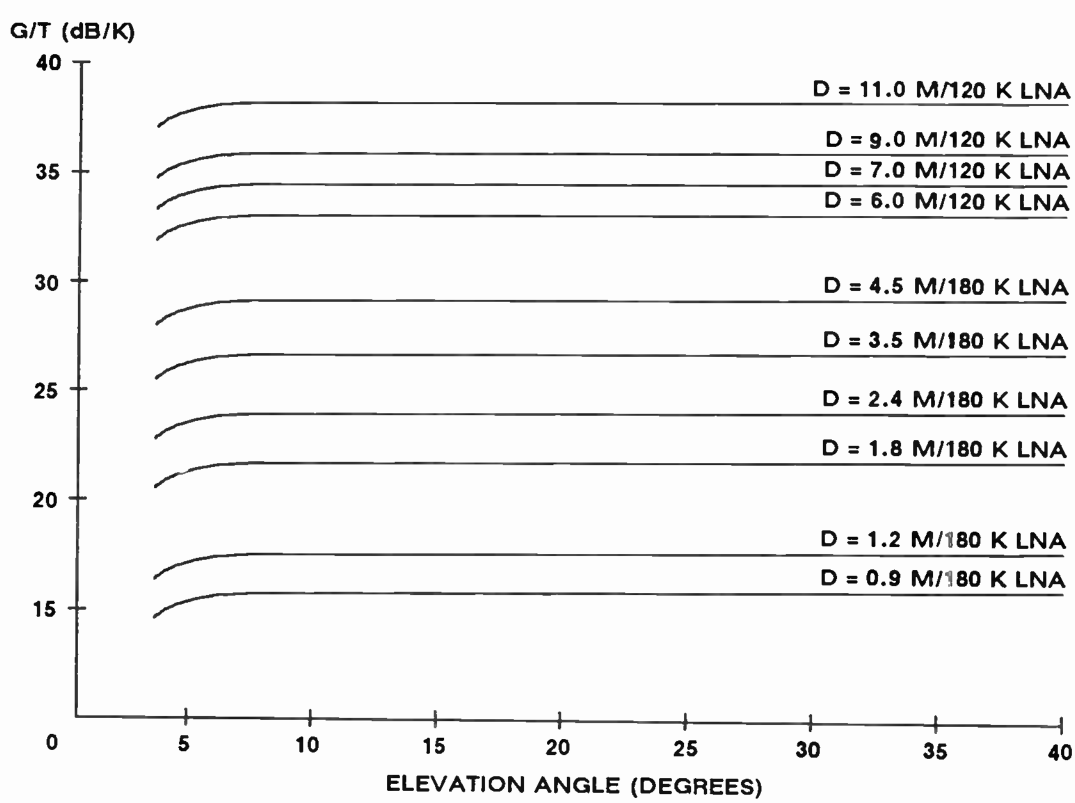
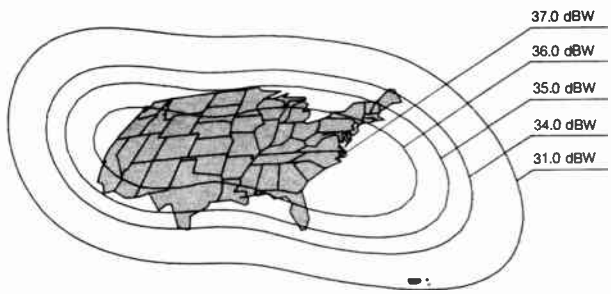
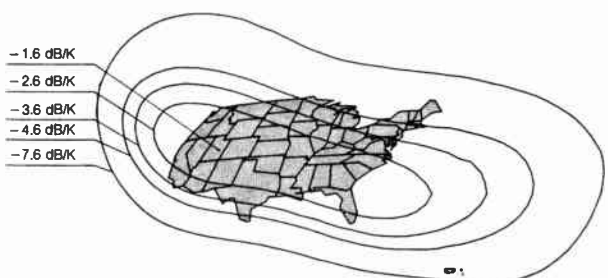


Figure 6.11-11b. Typical Ku-band g/T system performance versus elevation angle for different commonly used antenna diameters.



	EIRP (dBW)
ATLANTA	36.7
BOSTON	35.4
CHICAGO	36.4
DALLAS	36.0
HOUSTON	35.3
LOS ANGELES	35.3
NEW YORK	36.1
ORLANDO	35.9
SAN FRANCISCO	35.9
SAN JUAN	31.6
SEATTLE	36.0

Figure 6.11-12a. C-band satellite EIRP footprint.



	G/T (dB/K)
ATLANTA	−1.9
BOSTON	−4.8
CHICAGO	−3.5
DALLAS	−2.2
HOUSTON	−3.1
LOS ANGELES	−3.5
NEW YORK	−4.4
ORLANDO	−2.2
SAN FRANCISCO	−2.5
SAN JUAN	−6.8
SEATTLE	−2.4

Note SPACENET 36-MHz C-Band transponder saturation flux density (SFD) is ground-commandable to a nominal value of −86 or −80 dBW/m² corresponding to a G/T level of −5 dB/K

Figure 6.11-12b. C-band satellite G/T footprint.

Downlink C/N

Likewise the downlink $(C/N_o)_d$ can be calculated by:

$$(C/N_o)_d = EIRP_d - L_d + (G/T)_{e.s} + 228.6 \qquad [21]$$

and

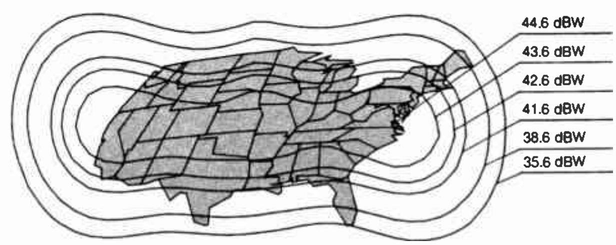
$$EIRP_d = EIRP_s - BO_o \qquad [22]$$

where: $EIRP_d$ = Downlink EIRP (dBW)
$EIRP_s$ = Saturated EIRP (dBW)
L_d = Downlink space loss (dB)
$(G/T)_{e.s}$ = Earth station G/T (dB/K)
BO_o = Transponder output Back-Off (dB)

It is important to note that Equations 20 and 22 are related by the non-linear power transfer function of the transponder, therefore, for transponder operation below saturation the input and output relationship needs to be resolved graphically with the aid of Figure 6.11-15 or its equivalent.

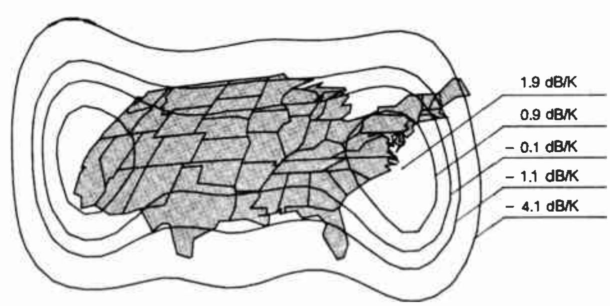
Once uplink and downlink noise contributions are determined the composite link performance in terms of total carrier to noise density ratio $(C/N_o)_t$ can be readily obtained by simple noise power addition since the uplink and downlink contributions are incoherent. This yields:

$$(C/N_o)_t = \{(C/N_o)_u^{-1} + (C/N_o)_d^{-1}\}^{-1} \qquad [23]$$



	EIRP (dBW)
ATLANTA	43.2
BOSTON	41.7
CHICAGO	42.9
DALLAS	43.2
HOUSTON	41.3
LOS ANGELES	43.0
NEW YORK	43.4
ORLANDO	39.8
SAN FRANCISCO	44.1
SEATTLE	42.0

Figure 6.11-13a. Ku-band satellite EIRP footprint.



	G/T (dB/K)
ATLANTA	−.1
BOSTON	−1.1
CHICAGO	.2
DALLAS	−.2
HOUSTON	−1.0
LOS ANGELES	1.2
NEW YORK	.7
ORLANDO	−1.4
SAN FRANCISCO	1.5
SEATTLE	−.3

Note SPACENET Ku-Band transponder saturation flux density (SFD) is ground-commandable to a nominal value of −86, −80 or −74 dBW/m² corresponding to a G/T level of −2 dB/K

Figure 6.11-13b. Ku-band satellite G/T footprint.

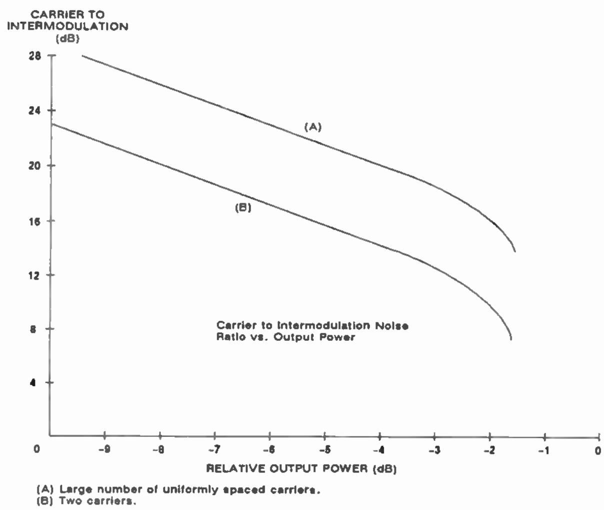


Figure 6.11-14. Output power normalized to single carrier saturation point.

This equation represents a simplified situation in that only thermal noise is added to the carriers. In actuality there are other sources of perturbations and interference; transponder non-linearity is the cause of some of the more important ones. As shown in Figure 6.11-15, operating the transponder near maximum power for better efficiency implies that compression, due to the instantaneous non-linear transfer characteristic of the amplifier relative to the signal amplitude, becomes more significant. Under this condition, when more than one frequency is amplified, interaction between the signals occurs and consequently a spectrum of spurious frequencies or intermodulation products is generated.

Particularly, the so called third order intermodulation product of the form $(2f_1 - f_2)$, a consequence of the third order non-linearity of the transponder, constitutes a significant interfering signal since it is the largest product and it falls in the same operating bandwidth of the information signal. Figure 6.11-14 shows how carrier-to-intermodulation ratio varies as a very sensitive function of the transponder operating output back-off.

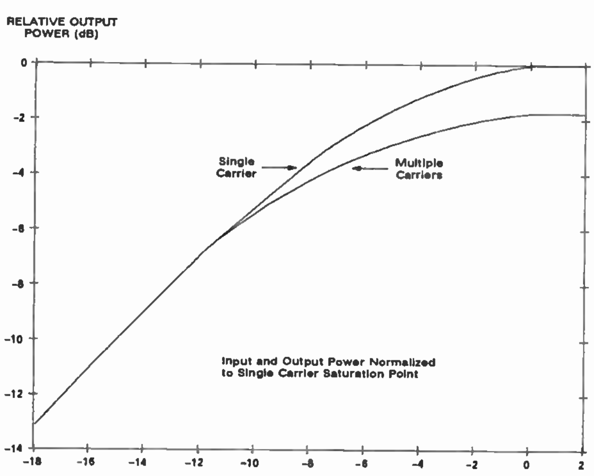


Figure 6.11-15. TWT power transfer characteristics.

System C/N

Figure 6.11-16 depicts a complete satellite link. Other sources of interference have been added, such as uplink interference due to off beam radiation from other earth stations and uplink cross-polarization isolation, combined, represented by the quantity $(C/I_o)_u$. Similarly in the downlink the quantity $(C/I_o)_d$ represents the combined effects of downlink cross-polarization isolation and adjacent satellite interference. When all these terms are considered the total link $(C/N_o)_t$ can be calculated by:

$$(C/N_o)_t = \{(C/N_o)_u^{-1} + (C/I_o)_u^{-1} + (C/N_o)_d^{-1} + (C/I_o)_d^{-1}\}^{-1} \quad [24]$$

Figure 6.11-17 shows the typical interaction of the different terms in Equation 24 as a function of transponder input back-off and in the presence of thermal and intermodulation noise. The total $(C/N_o)_t$ can be maximized by reducing the transponder input drive and adjusting the transponder gain. Backing off the TWT reduces $(C/N_o)_u$ and also $(C/N_o)_d$ (through the input/output relationship of the transponder), but as $(C/N_o)_I$ increases rapidly when the input drive is reduced, an optimum value of $(C/N_o)_t$ is obtained at a specific back-off level. Interference noise can be kept down by proper antenna design, transponder sensitivity and frequency coordination of satellite services.

Rain Effects on System Performance

Rain is the dominant factor in satellite propagation for frequencies above 10 GHz. Rain propagation has been studied intensively [10] since the late 1960s and only a brief discussion will be presented here. Due to the basic interaction of electromagnetic waves with water in liquid form, raindrops cause absorption, scattering and depolarization phenomena. Absorption and

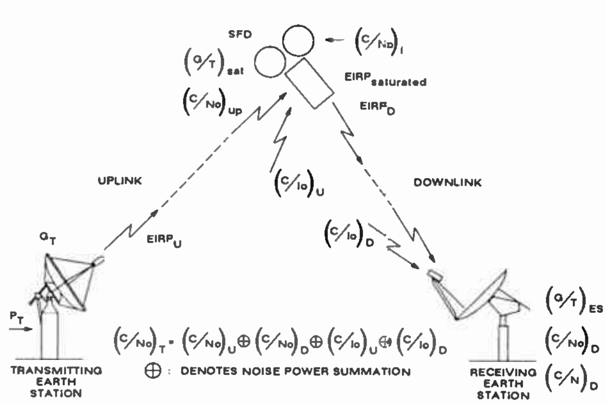


Figure 6.11-16. Satellite link model.

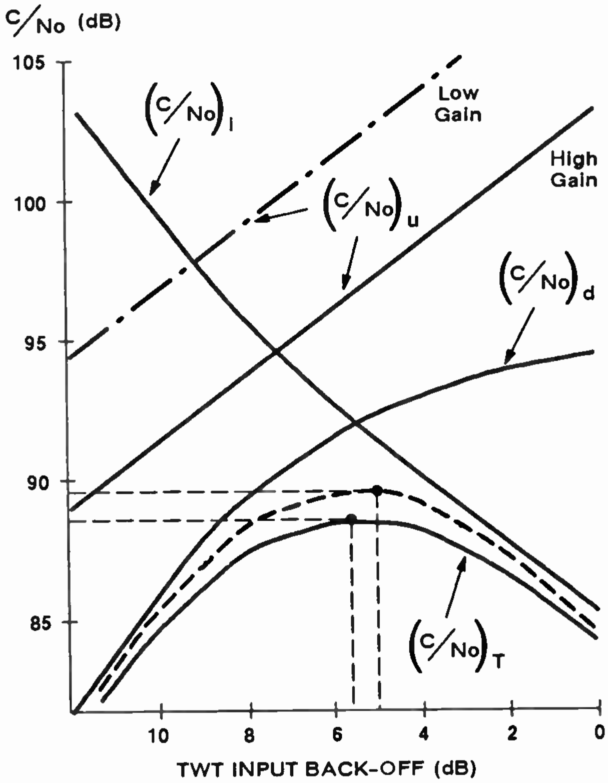


Figure 6.11-17. Optimum TWT operation.

scattering result in signal attenuation and an increase in sky noise temperature, with the consequent degradation of the received C/N_o. Depolarization has an effect on dual polarization systems and creates interference between cross-polarized signals.

Signal Attenuation

The amount of attenuation depends fundamentally on the rain intensity or rain rate and the signal path

Figure 6.11-19. Rain rate climate regions for Europe. (From NASA Propagation Effects Handbook for Satellite System Design, ORI TR 1679.)

length in rain. Rainfall data are available for most parts of the world; different types of climates have been defined and boundaries of their regions identified. Figures 6.11-18 through 20 show the NASA rain rate climate regions. The long term behavior of rain is described by the cumulative probability distribution or exceedence curve. This gives the percentage of time that the rain rate exceeds a given value. Table 6.11-2 gives the rain-rate distribution values versus per cent of year for the various rain climate regions of Figures 6.11-18 through Figure 21. Figures 6.11-21 (a) and (b) plot the rain rate cumulative probability distributions for the regions presented on the previous maps.

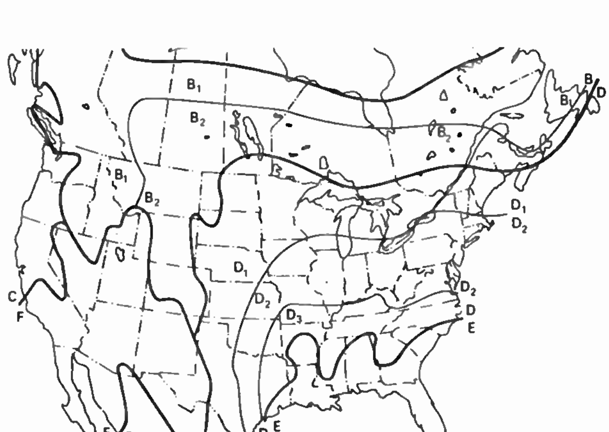


Figure 6.11-18. Rain rate climate regions for the continental United States showing the subdivision of Region D. (From NASA Propagation Effects Handbook for Satellite System Design, ORI TR 1679.)

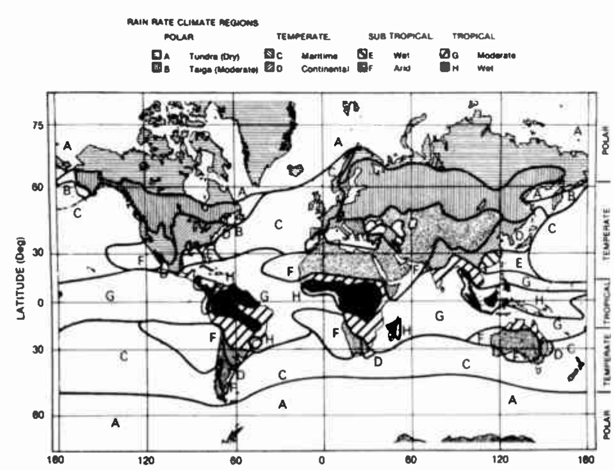


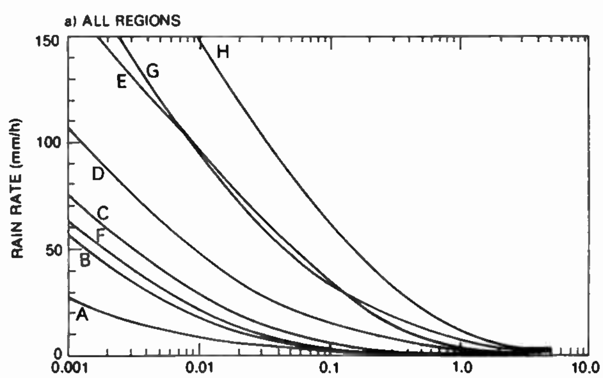
Figure 6.11-20. Global rain rate climate regions, including the ocean areas. (From NASA Propagation Effects Handbook for Satellite System Design, ORI TR 1679.)

Table 6.11-2
Point rain rate distribution values (millimeters per hour) vs. percent of the year rain rate is exceeded.

Percent of Year	Rain Climate Region A	B	C	D_1	D_2	D_3	E	F	G	H	Minutes Per Year	Hours Per Year
0.001	28.0	54.0	80.0	90.01	02.0	127.0	164.0	66.01	29.0	2512.0	5.3	0.09
0.002	24.0	40.0	62.0	72.0	86.0	107.0	144.0	51.0	109.0	220.0	10.5	0.18
0.005	19.0	26.0	41.0	50.0	64.0	81.0	117.0	34.0	85.0	178.0	26.0	0.44
0.01	15.0	19.0	28.0	37.0	49.0	63.0	98.0	23.0	67.0	147.0	53.0	0.88
0.02	12.0	14.0	18.0	27.0	35.0	48.0	77.0	14.0	51.0	115.0	105.0	1.75
0.05	8.0	9.5	11.0	16.0	22.0	31.0	52.0	8.0	33.0	77.0	263.0	4.38
0.1	6.5	6.8	7.2	11.0	15.0	22.0	35.0	5.5	22.0	51.0	526.0	8.77
0.2	4.0	4.8	4.8	7.5	9.5	14.0	21.0	3.8	14.0	31.0	1052.0	17.50
0.5	2.5	2.7	2.8	4.0	5.2	7.0	8.5	2.4	7.0	13.0	2630.0	43.80
1.0	1.7	1.8	1.9	2.2	3.0	4.0	4.0	1.7	3.7	6.4	5260.0	87.66
2.0	1.1	.2	1.2	1.3	1.8	2.5	2.0	1.1	.6	2.8	10520.0	175.30

Source: NASA Propagation Effects Handbook for Satellite Systems Design, ORITR 1679

The calculation of the rain attenuation involves two basic steps. The first step is to determine the rain rate in mm/hr as a function of the cumulative probability of occurrence. This probability will be defined by the grade of service or availability of the link to be provided. The second step consists of the calculation of the actual rain attenuation associated with the rain rate that was exceeded with such probability.



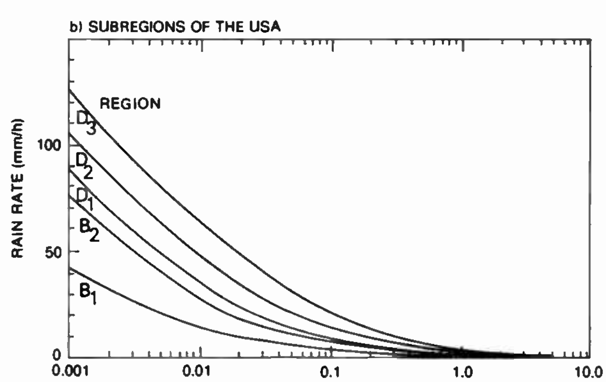


Figure 6.11-21. Rain rate cumulative probability distributions for the regions presented on the previous maps. (reprinted from Louis J. Ippolito, R.D. Kaul, and R.G. Wallace, Propagation Effects Handbook for Satellite Systems Design [NASA Reference Publication 1082(03)], National Aeronautics and Space Administration, Washington, DC, June 1983. Courtesy of NASA.)

The attenuation per unit of length (specific attenuation), λ_r (dB/Km), is tied to the rain rate R (mm/hr), by the empirically derived relationship:

$$\lambda_r = a(f)R^{b(f)} \quad \text{(dB/km)} \qquad [25]$$

where $a(f)$ and $b(f)$ are frequency dependent coefficients. For the frequency range between 8.5 and 25 GHz, Equation 25 becomes:

$$\lambda_r = 4.21*10^{-5} f^{2.42}*1.41*f^{-0.0779} R \quad \text{(dB/km)} \qquad [26]$$

The attenuation per unit length is heavily frequency dependent, Figure 6.11-22 shows frequency dependence of λ_r for various rain rates.

Introducing the concept of equivalent path length, $L_e(R)$, the total rain attenuation in decibels is simply:

$$Ar = \lambda_r L_e(R) \quad \text{(dB)} \qquad [27]$$

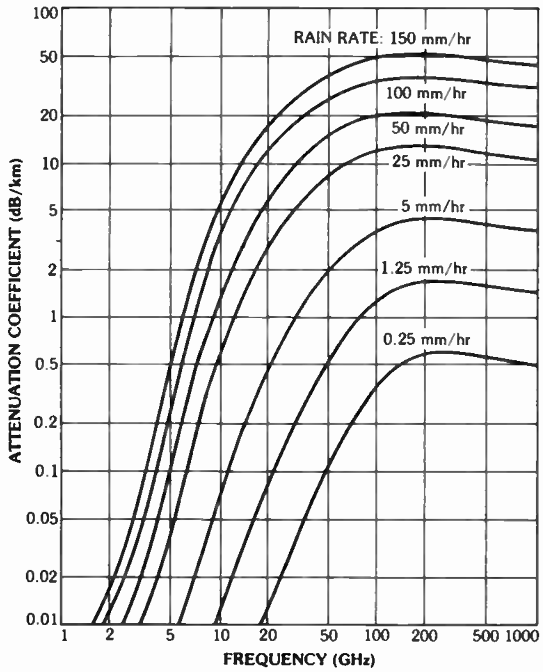
Equivalent path length is primarily determined by the height of the freezing level or 0° isotherm, which depends on latitude, season and rain rate, the cosecant of the elevation angle and site altitude. For latitudes within ±30°, the freezing level is at 4.8 km. Curves of equivalent path lengths versus elevation angle and for different rain rates are shown in Figure 6.11-23.

The rain attenuation is required to be added to the satellite link as a margin to allow the specified availability under fading conditions. Figures 6.11-24(a) and (b) show typical rain attenuations versus rainfall rate in the transmit and receive Ku-bands for different elevations angles.

Noise Contribution

In addition to the attenuation, rain also degrades the performance of a satellite link by increasing the earth station antenna noise temperature. In clear weather the antenna sees the cold background of space, but in rain it receives thermal radiation from the raindrops. The increase in antenna noise temperature due to rain, T_r, may be estimated by:

$$T_r = 280(1 - 10^{-A/10}) \quad \text{(K)} \qquad [28]$$



Raindrop size distribution: Laws and Parsons, 1943
Terminal velocity of raindrops: Gunn and Kinzer, 1949
Dielectric constant of water at 20°C: Ray, 1972

Figure 6.11-22. Attenuation per unit length versus frequency and rain rate. (From K. Miya, ed., Satellite Communications Technology. Tokyo: KDD Engineering and Consulting, Inc., 1982.)

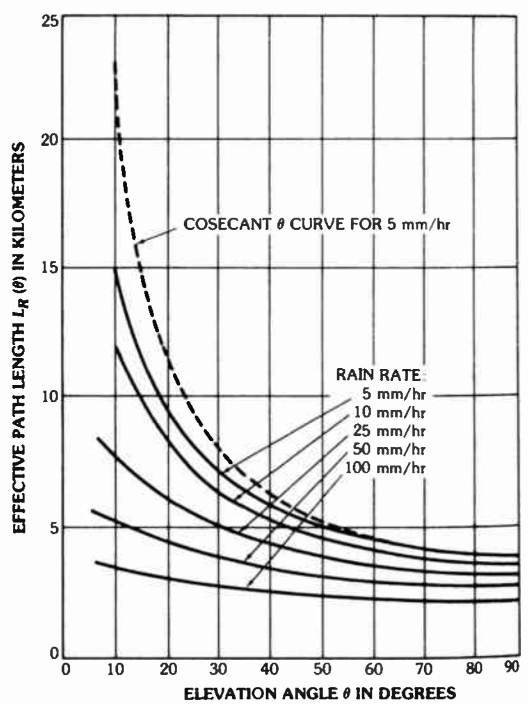


Figure 6.11-23. Equivalent path length versus rain rate and elevation angle. (from K. Miya, ed., Satellite Communications Technology. Tokyo: KDD Engineering and Consulting, Inc., 1982.)

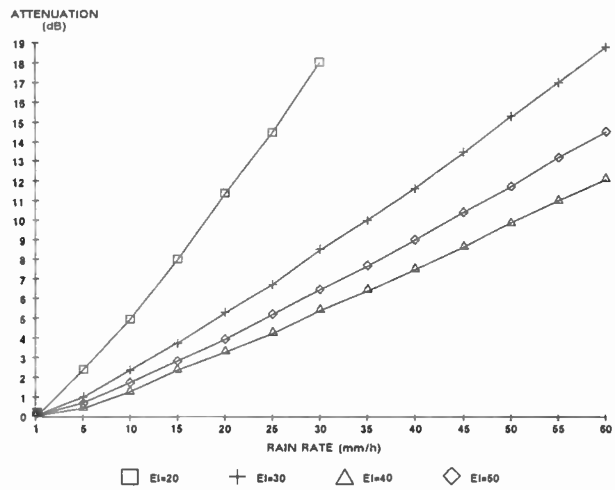


Figure 6.11-24a. Rain attenuation (11.95 GHz) 4.8 km zero-degree isotherm.

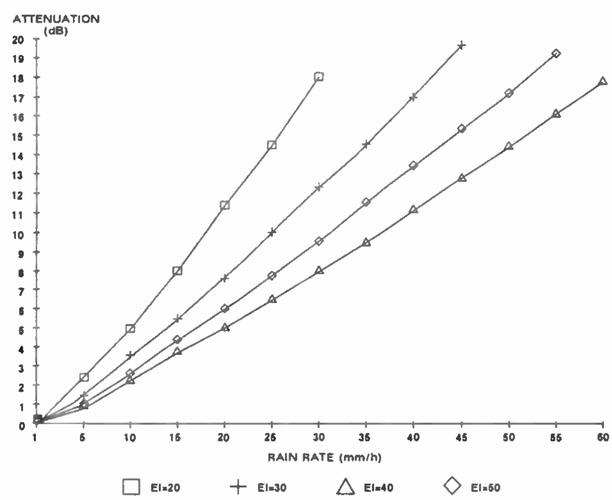


Figure 6.11-24b. Rain attenuation (14.25 GHz) 4.8 km zero-degree isotherm.

where A is the rain attenuation in decibels. Figure 6.11-25 shows the impact of the rain contribution of noise temperature on the normal clear sky *G/T* for different clear sky system temperatures. The *G/T* degradation corresponding to the rain attenuation for the stipulated link availability also must be added to the satellite downlink. This is to provide sufficient margin to compensate for the combined rain effect of signal attenuation and noise increase.

The allocations of rain fade margins in the uplink and downlink can be done independently, corresponding to specific availability requirements of the uplink or downlink and consistently with the availability requirement of the total link. The assumption is that, due to the localized nature of the rain fades, the uplink fade and downlink fade can be considered as two statistically independent processes. Therefore, total link availability can be obtained as the reciprocal of the

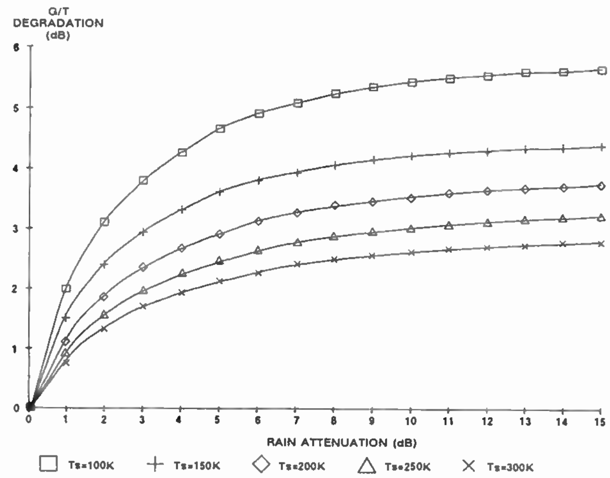


Figure 6.11-25. G/T rain degradation.

summation of the uplink and downlink outages calculated as if they occurred independently and one at a time.

Example of System Link Calculation

Table 6.11-3 shows a typical satellite link budget for a video uplink and downlink for Ku-band operation where a 5 dB uplink power control has been applied to mitigate the effects of rain fade in the uplink. A C-band link budget would contain similar terms but the uplink power control to mitigate the effects of rain would not be necessary.

Baseband Performance Link Analysis

As stated previously, the overall quality of the delivered baseband signal can be expressed in analog systems by the S/N or in digital systems by the BER. Other performance parameters such as phase and frequency

Table 6.11-3
Link Budget for Typical Ku-band Video Satellite Link

Satellite		Spacenet-II			
Beam Type		Conus			
Type of Service		FM/Video			
Transmit/Receive Connectivity		7.0/4.5 meter			
Occupied Bandwidth per Carrier		30.0 MHz			
Available Bandwidth per Carrier		36.0 MHz			
Transponder Bandwidth		72.0 MHz			
Parameter			Values		
I. UPLINK NOISE		Clear Sky	Uplink Fade	Downlink Fade	Units
Earth Station EIRP per Carrier		75.0	80.0	75.0	dBW
Pointing Losses		0.5	0.5	0.5	dB
Path Loss		207.0	207.0	207.0	dB
Isotropic Antenna Area		44.5	44.5	44.5	dBW/m²
Saturation Flux Density		−81.0	−81.0	−81.0	dBW/m²
Rain Attenuation		0.0	6.0	0.0	dB
G/T including Footprint Advantage		−1.1	−1.1	−1.1	dB/K
Input Back-off per Carrier		7.0	8.0	7.0	dB
Uplink Thermal C/N		20.2	19.2	20.2	dB
Co-channel Interference		27.0	24.0	27.0	dB
Off Beam Emissions Interference		26.0	25.0	26.0	dB
Total Uplink C/(N + I)		18.5	17.2	18.5	dB
UPLINK AVAILABILITY	99.99%				
II. INTERMODULATION NOISE		20.0	19.0	20.0	dB
III. DOWNLINK NOISE					
Satellite Saturation EIRP		43.0	43.0	43.0	dBW
Transponder Output Back-off/Carrier		4.5	6.5	4.5	dB
EIRP per Carrier		38.5	36.5	38.5	dBW
Path Loss		206.0	206.0	206.0	dB
Rain Attenuation		0.0	0.0	3.0	dB
Pointing Losses		0.5	0.5	0.5	dB
Earth Station G/T		29.5	29.5	29.5	dB/K
G/T Degradation		0.0	0.0	2.2	dB
Downlink Thermal C/N		15.3	13.3	10.1	dB
Co-channel Interference		27.0	25.0	19.8	dB
Total Downlink C/(N + I)		15.0	13.0	9.7	dB
DOWNLINK AVAILABILITY	99.85%				
IV. TOTAL C/(N + I) NOISE					
Total C/(N + I)		12.6	10.9	8.8	dB
Occupied Bandwidth per Carrier		74.8	74.8	74.8	dB-Hz
$C/(N_o + I_o)$ Total		87.4	85.7	83.6	dB-Hz
Required C/N_o		83.5	83.5	83.5	dB-Hz
Margin		3.9	2.2	0.1	dB
TOTAL LINK AVAILABILITY	99.75%				

linearity and intersymbol interference, can be utilized to fully characterize quality as is common in any transmission system. However, since satellites are inherently wide band repeaters, S/N and BER are the most sensitive performance factors that depend on link operational parameters.

The C/N_o versus S/N performance of different modulation schemes can be characterized by rather simple mathematical equations. The most common analog modulation technique in satellite transmission is frequency modulation (FM). Because of its simplicity and the low cost of the receivers and demodulators, FM is widely utilized in the transmission of television signals. Equation 29 allows the computation of the S/N as a function of C/N_o and modulation parameters.

$$(S/N)_w = C/N_o + 10\log(12\Delta F^2/B_v^3) + W \quad [29]$$

or in terms of carrier-to-total-noise power ratio in a bandwidth B:

$$(S/N)_w = C/N + 10\log(12\Delta F^2/B_v^3) + 10\log(B/B_v) + W \quad [30]$$

where:

$(S/N)_w$ = Weighted Signal-to-noise Ratio (dB)
C/N_o = Carrier-to-noise density ratio (dB)
ΔF = Peak composite video deviation (MHz)
B = IF predetection noise bandwidth (MHz)
B_v = Video filter bandwidth (MHz)
W = Deemphasis and weighting improvement (dB)

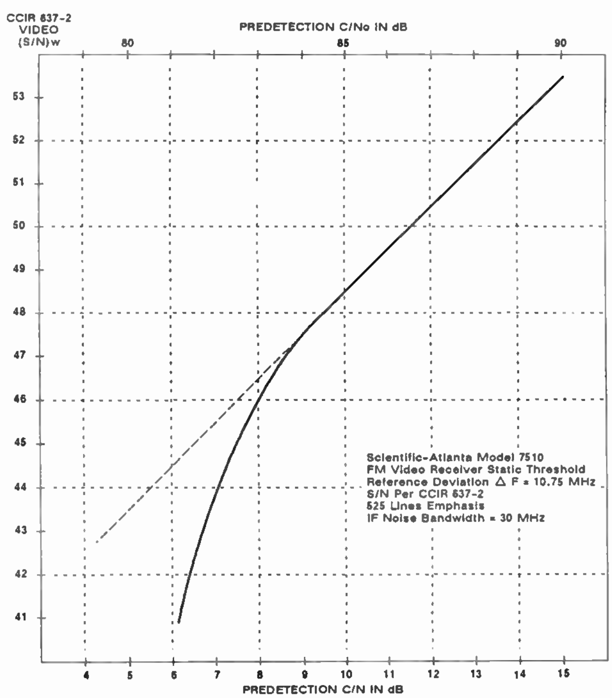


Figure 6.11-26. Signal-to-noise ratio performance of FM video demodulator vs. C/N and C/N_o.

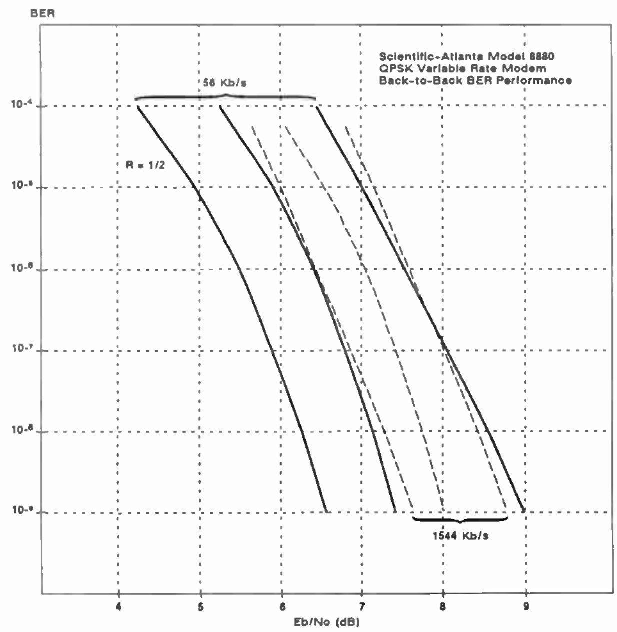


Figure 6.11-27. BER modem performance vs. bit energy over noise density ratio.

For NTSC and 30 MHz bandwidth these parameters are typically: ΔF = 10.75 MHz, B = 28 MHz, B_v = 4.2 MHz and W = 13.8 dB. Figure 6.11-26 shows the result of Equation 29 with the previous transmission parameters at high levels of C/N. The departure from a linear relationship at low values of C/N is not predicted by Equation 29 but represents the actual performance of typical FM demodulator. This phenomenon is known as the *threshold effect.*

In digital satellite systems the most common modulation technique is phase shift keying (PSK). Variations of this technique are known as binary PSK or BPSK in the case that logic symbols zeros and ones are mapped into RF signals 180° apart in phase; and as Quadrature PSK or QPSK when the phases are 90° apart. In recent years 8PSK, using 40° phases, has also come into widespread use. The BER performance of these systems is evaluated as a function of the energy per bit of information transmitted (E_b) versus noise density, E_b/N_o. Sophisticated digital coding and decoding techniques exist that, by adding error control bits to the information data stream, allow substantial improvements in BER that can translate into transmit power reductions of up to 5 dB. The ratio between the uncoded data rate and the coded one is called coding rate (R). Figure 6.11-27 shows the performance of a typical QPSK modulator and demodulator for different data and coding rates.

Earth Station Block Diagram

The block diagram of Figure 6.11-28 depicts an earth station capable of providing uplink services for both encrypted video and data in the vertical and hori-

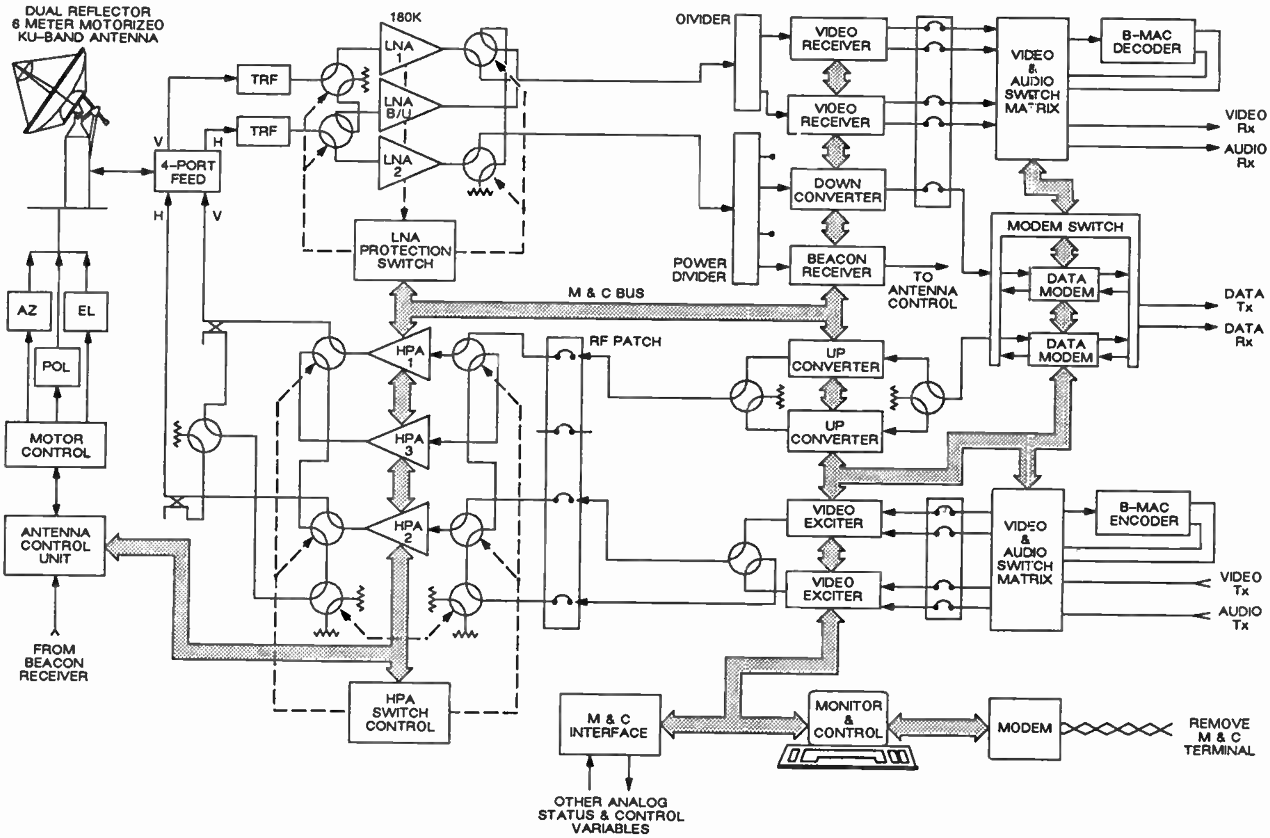


Figure 6.11-28. Encrypted video and data broadcast earth station block diagram.

zontal polarizations simultaneously. All subsystems are redundant for maximum reliability. The backup video exciter and HPA are also capable of occasional video uplinks through transmit couplers in either polarization. A computer based monitor and control system, by means of a serial control bus, offers centralized operation of the complete earth station with the ability of monitoring all status and controlling all variable parameters of every subsystem from the local or remote terminals.

Figure 6.11-29 shows the block diagram of the corresponding dual polarization receive only terminal. This low cost earth station, with an L-band (950 MHz to 1450 MHz) inter-facility link (IFL) can provide simultaneous reception of encrypted video and data.

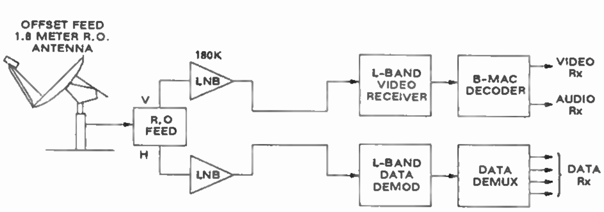


Figure 6.11-29. Receive-only earth station for encrypted video and data broadcast.

Interference Analysis

The consideration of interference in a satellite communication system is important, not only from the standpoint of the desired satellite signal being interfered with, but also as it applies to the satellite signals generating interference into other systems. It is mandatory for a proposed satellite transmit system in the United States to submit a coordination filing that includes an interference analysis to the FCC. This analysis must show the impact of the proposed system on existing operational systems and must satisfy the allowable interference requirements of the FCC. Coordination for receive only systems is not mandatory and is only necessary when a C-band system desires interference protection from future terrestrial systems.

In August 1983, the FCC finalized a new satellite orbital assignment plan based on a frequency and polarization plan to allow satellite spacing to be reduced from the previous 4° to 2° with an interim average spacing of 2.5° at C-band. The implementation of this plan depended on several important technical achievements including:

- The adoption of a [29 − 25 log(θ)] dBi peak sidelobe envelope for angles off boresight between 1 and 7°
- Frequency reuse for all satellites

- Adjacent satellite, same frequency transponders being orthogonally polarized
- Homogeneity of satellite EIRP and saturation flux density characteristics for minimum spacing

Model for Interference Analysis

The interference analysis presented in the following sections is based on models for the antenna characteristics, the spectral characteristics for the desired and interfering carriers and assumes the final implementation of a uniform 2° spaced satellite environment about the geostationary orbital arc.

Antenna Characteristics

The primary characteristic of the antenna which affects the interference analysis is the angular discrimination, the gain differential between the on-axis gain and the gain of an off-axis angle for the interfering source. For this analysis, the copolarized radiation patterns of the assumed earth station antennas are characterized by the $[29 - 25 \log(\theta)]$ dBi envelope. The cross-polarized radiation patterns are characterized by $[19 - 25 \log(\theta)]$ dBi envelope (see Figure 6.11-30). In actuality, the antenna radiation sidelobe may fall below or above these reference envelopes by some predetermined acceptable level. Any sidelobes that are below the reference envelope and at the appropriate pointing angles of adjacent satellites would reduce the interference and, conversely, any sidelobes above the envelope pointing at adjacent satellites would increase the interference. The cross-polarization discrimination of 10 dB is assumed to apply for clear sky conditions. It should be pointed out that during periods of heavy rain, the polarization of the incoming/outgoing signals may be affected such that the full 10 dB is not realized.

Satellite Characteristics

The analysis that follows is based on a satellite deployment model with cofrequency transponders on adjacent satellites being cross-polarized with each other. The calculations are performed with this model

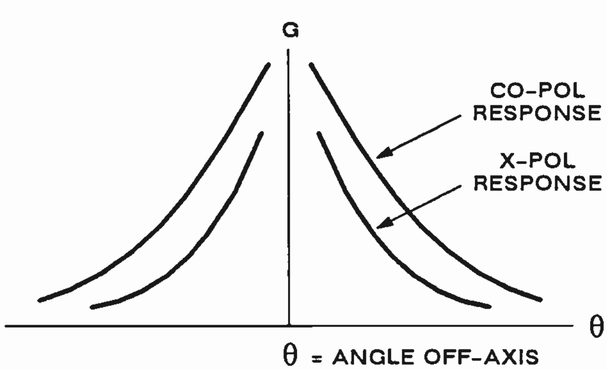


Figure 6.11-30. Earth station antenna radiation characteristics.

to demonstrate the expected results in a uniform satellite environment. Three cases of this model are examined:

1. A homogeneous model in which interfering and desired satellites have the same saturation flux density and radiated EIRP. (The radiation patterns yield the same signal strength at any given location on the ground.)
2. A model in which the interfering satellite EIRP exceeds the desired satellite EIRP by 2 dB.
3. A model in which the interfering satellite EIRP exceeds the desired satellite EIRP by 4 dB.

Note that the spacecraft antennas are assumed to have a minimum cross-polarization discrimination of 35 dB.

EIRP is a very important consideration. Antenna and transponder characteristics are such that their initial EIRP contours on the earth's surface are not identical. Differences in the initial EIRP contours and differences in transponder aging must be considered in a practical system. An orbital spacing plan that is predicated on differential EIRPs of less than 2 dB represents an impractical burden, both on the satellite manufacturers, and on the FCC in assuring compliance with a more stringent specification. Therefore, it is suggested that the calculations for the second case (2 dB variations in EIRPs) be taken as representing a practical case.

The calculations will also be based on geosynchronous rather than topocentric angles and do not include stationkeeping inaccuracies. An average topocentric angle for the contiguous United States (CONUS) can be estimated by multiplying the geocentric angle by 1.08.

Sources of Interference

Interference into a geostationary satellite communication system can originate from several sources, including the following:

- Adjacent satellite signals
- Internal cross-polarization signals (half transponder frequency offsets)
- Terrestrial microwave signals (does not apply for Ku-band or Ka-band)

These three are analyzed separately in the following paragraphs and then combined to determine the total interference into the system.

Adjacent Satellite Interference

Interference from adjacent satellites occurs in two ways: uplink interference from earth stations transmitting to adjacent satellites and downlink interference from adjacent satellite transmission into the desired earth station. The interference in both the uplink and downlink consists of many signals (23 of 24 channels for a fully loaded frequency reuse satellite), but it is primarily caused by the co-frequency channels/or transponders and the two half transponder bandwidth offset-channels in a frequency reuse system, Figure 6.11-31. The particular interferers for the C-band example system are the following:

a) The co-frequency, cross-polarized channel on the first adjacent satellite on each side.
b) The two 20 MHz offset-frequency, co-polarized channels on the first adjacent satellite on either side.
c) The co-frequency, co-polarized channel on the second adjacent satellite on either side.
d) The two 20 MHz offset-frequency, cross-polarized channels on the second adjacent satellite on either side.

The contribution to interference from satellites at orbital positions greater than 4° from the desired satellite tends to be noise-like in that it is the result of a number of small, relatively non-coherent signals.

The equations for calculation of the adjacent satellite interference are given below:

$$(C/I)_u = (EIRP)_{es} - \sum_{i=1}^{N} \oplus \{(EIRP)_i - [G_i - G(\theta_i)] + F_i + P_i\} \quad [31]$$

where:

$\sum_{i=1}^{N} \oplus$ = Series power summation
$(EIRP)_{es}$ = Earth station radiated power in dBW
$(EIRP)_i$ = Effective radiated power of interfering earth station in dBW
G_i = Peak gain of the interfering earth station in dBi
$G(\theta)_i$ = Gain of the interfering earth station in direction (θ) in dBi
F_i = Frequency discrimination factor for ith earth station
P_i = Polarization discrimination factor for ith earth station

$$(C/I)_d = (EIRP)_{sat} + G_{es} - \sum_{i=1}^{N} \oplus \{(EIRP)_i + G_{es}(\theta_i) + F_i + P_i\} \quad [32]$$

where

$(EIRP)_{sat}$ = Effective radiated power of satellite in the direction of the receive earth station in dBW
G_{es} = Gain of the receive earth station in dBi
$G_{es}(\theta_i)$ = Gain of the receive earth station in the direction θ_i in dBi
F_i = Frequency discrimination factor for the ith earth station
P_i = Polarization discrimination factor for the ith earth station
N = Number of transponders considered ($N \geq 3$)
M = Number of adjacent satellites considered ($M \geq 4$, typically 8, 4 on each side)

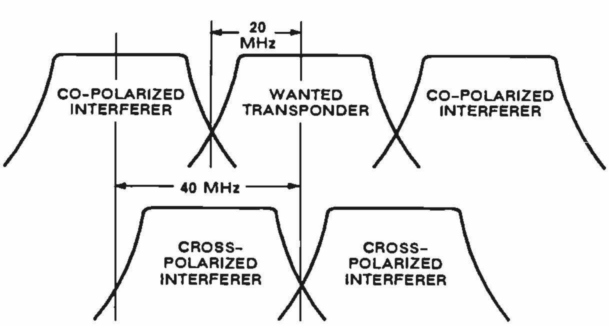


Figure 6.11-31. Frequency reuse transponder plan.

The total adjacent satellite interference is then calculated by combining the uplink and downlink contributions in a power summation manner.

$$(C/I)_{adj.sat} = (C/I)_u \oplus (C/I)_d \quad [33]$$

The polarization discrimination factor in the previous equations is the system discrimination rather than that of the receive or transmit antenna alone. A well designed dual linearly polarized antenna can achieve excellent cross-polarization discrimination on or near the main beam axis (greater than 35 dB or 40 dB relative to the copolarized energy) and reasonable rejection of the cross-polarized signals in the close-in sidelobe regions. The adjacent satellite signals are received through the sidelobes of the earth station, and the $[19 - 25 \log(\theta)]$ envelope is assumed in the analysis. This assumption, rather than being conservative, may be optimistic when one considers the interactions of the ionosphere and atmosphere on the transmitted signals (from the earth station and/or the satellites) and the polarization angle alignment between satellites. A more conservative analysis may assume a slightly reduced discrimination of perhaps $[21 - 25 \log(\theta)]$ for the cross-polarized sidelobe energy in the off-axis regions.

The frequency discrimination factor is related to the spectra of the desired and undesired signals. This factor can range from 0 to 12 dB depending upon the interfering power from the different services. For example, the F_i term, where an FM/TV signal is interfering with a FM/TV signal occupying the same bandwidth, would be 0 dB. For a 20-MHZ offset-frequency FM/TV signal, with a 36 MHz bandwidth, F_i could range from 3 to 15 dB depending on the characteristics of the video signals. It is suggested that for typical FM/TV a value 6.5 to 8.0 dB is realistic. The 8.0 dB value is suggested by the FCC, but a conservative value of 6.5 dB will be used in the example analysis.

Internal Interference

The internal interference in a satellite system is primarily due to the two adjacent 20-MHz offset-frequency channels. The interfering power from different services has been calculated by convolving the power spectra of the individual services. This data, taken together with the appropriate polarization discrimination term, determines the amount of interference and is given in Table 6.11-4. The polarization discrimination

Table 6.11-4
Summary of Internal Interference for TV/FM Service.

I.			Uplink Pol Disc		Downlink Pol Disc
Satellite			35.0 dB		35.0 dB
Ground Station			35.0		35.0
Faraday			35.0		29.0
Resultant			28.0 dB		26.0 dB
II. Atmospheric Effect—25° Elevation Angle					
% Time	Rain Rate		Uplink		Downlink
99.0	0.5 in/hr		30.5 dB		33.0 dB
99.9	1.5 in/hr		21.0		25.0
99.99	3.0 in/hr		16.5		20.0
III. Resultant Polarization Discrimination					
% Time			Uplink		Downlink
99.0			23.1 dB		22.8 dB
99.9			17.8		19.5
99.99			14.5		16.5
IV. Carrier-to-Interference TV/FM					
% Time	$(XPD)_u$	$(XPD)_d$	$(C/I)_u$	$(C/I)_d$	$(C/I)_s$
99.0	23.1	22.8	27.1	26.56	23.81
99.9	17.8	19.5	21.2	21.26	18.22
99.99	14.5	16.5	18.6	17.96	15.26

term, is dependent on the climatic conditions which are dealt with in Table 6.11-4 by a percent of time condition.

Terrestrial Interference

Terrestrial microwave carriers are centered on frequencies offset by 10-MHz from the satellite carriers. To analyze the effects of terrestrial carriers on the FM/TV system, it is necessary to determine the power level of the interfering signal and the spillover of the terrestrial carrier spectra into the passband of the receiver. The first factor involves site details, such as angular discrimination and distance to the interfering transmitter. The second factor can be computed from the spectral distribution projected for the terrestrial carrier and the filter characteristic of the receiver. For the purpose of this analysis, it is assumed the *C/I* due to terrestrial microwave is 25 dB.

Interference Analysis for FM/TV Service

The interference for the FM/TV service is based on the following parameters:

Parameter	Specification
Transponder EIRP	34 dBW
Antenna Size (diameter)	
Transmit	10 m
Receive	3 m, 4.5 m 7 m, or 10 m
Uplink IRP	80 dBW
Transmit Power	4.5 kW

Each of the antennas listed previously is presently used in FM/TV systems. Many are licensed and regulated and therefore protected from interference in certain respects. Many receive only earth stations are unlicensed and therefore not protected. The result of the analyses is presented in Figures 6.11-32 through 6.11-35. Each figure includes three cases of desired signal EIRP relative to the interfering signal EIRP and the effect of variable terrestrial interference is shown.

Sun Transits and Eclipses

Communication satellite systems experience predictable service interruptions involving the sun. A sun transit outage occurs when the pointing angles from a receiving earth station to a satellite and to the sun so nearly coincide that the additional noise power presented by the sun renders transmission unusable. A solar eclipse occurs when the earth shadows the sun from the satellite. The eclipse event is not as serious as the sun transit event since the satellite has battery backup systems to augment the solar primary power.

Daily sun transits of all geostationary satellites serving an earth station occur during one week in the spring and again in the fall. The exact dates depend primarily upon the latitude of the receiving earth station. The geometry and duration associated with a sun transit are controlled by the off-axis gain of the earth station antenna, the receiving system noise temperature, the solar noise power profile and the minimum acceptable

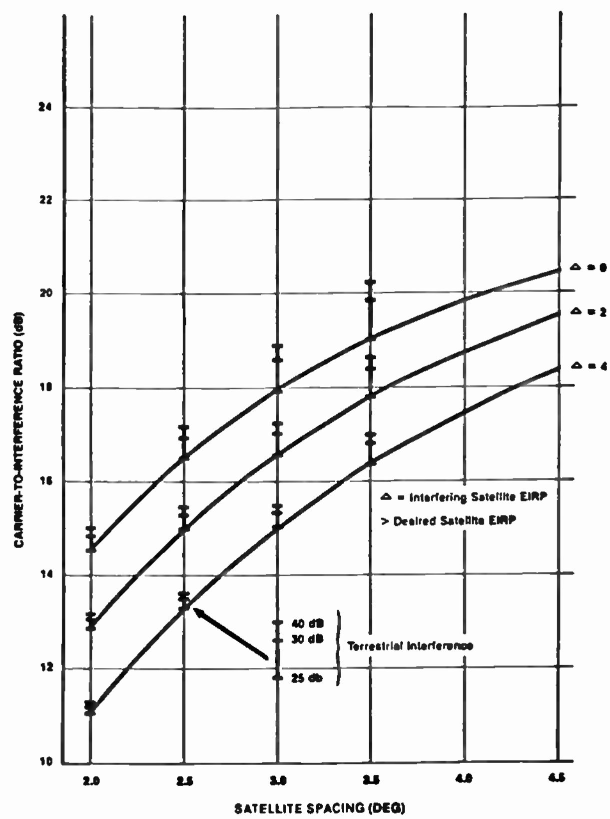


Figure 6.11-32. Carrier-to-interference ratio vs. satellite spacing for a 10-meter uplink and a 3-meter downlink. FX = 6.5 dB, PX = 10 dB. Co-Pol 29-25 Log θ envelope and adjacent satellite polarization interleaving.

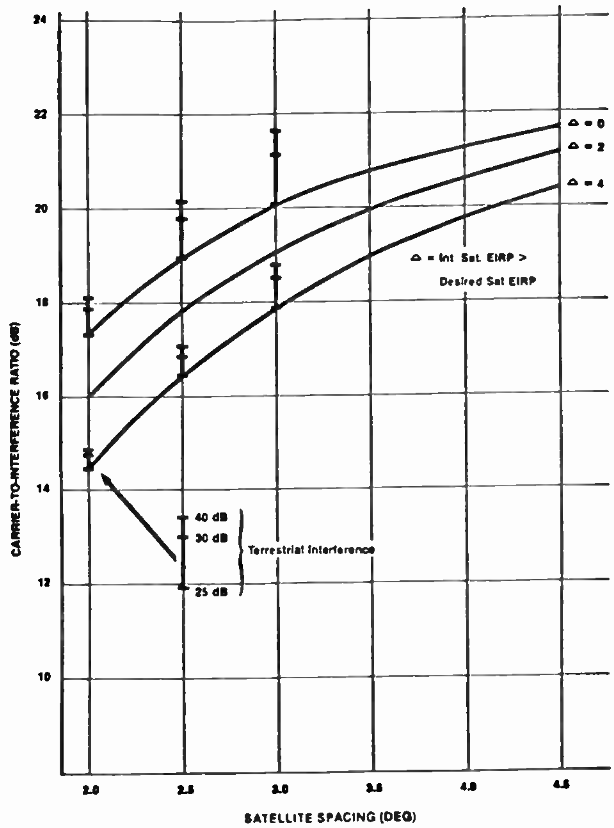


Figure 6.11-33. Carrier-to-interference ratio vs. satellite spacing for a 10-meter uplink and a 4.5-meter downlink. FX = 6.5 dB, PX = 10 dB. Co-Pol 29-25 Log θ envelope and adjacent satellite polarization interleaving.

S/N ratio. In late February or early March, short daily outages affect earth station systems situated near the United States-Canadian border. Two or three days later these systems experience maximum outages lasting five minutes or more, depending upon transmission parameters and permissible *S/N*. Outages at these earth station locations end after an additional two to three days and the sun transit outage paths progress southward at a rate of about 3° latitude per day. All outages affecting United States earth station antenna systems above north latitude 26° cease prior to mid-March. Conversely, in the fall the daily outages progress from south to north, affecting southern United States earth stations beginning October 1 and ending in the north about mid-October.

Eclipses of geostationary satellites can be expected for a total of about 90 evenings per year in the spring and fall. Eclipses occur near apparent midnight of the time zone at each satellite's longitude, beginning in late February or early March and ending mid-April. Fall events begin about September 1 and end mid-October. Eclipses of about 70 minutes duration occur on the dates of the spring and fall equinoxes. Communication satellites are provided with batteries to prevent circuit outages and to maintain pointing, attitude control, stationkeeping, telemetry and command capabilities during eclipses.

EQUIPMENT CHARACTERISTICS

An earth station system is made up of four major subsystems: the antenna subsystem, the transmitting subsystem, the receiving subsystem and the monitor and control subsystem.

Antenna

The antenna is one of the more important component parts since it provides the means of transmitting signals to the satellite and/or collecting the signal transmitted by the satellite. The antenna must not only provide the gain necessary to allow proper transmission and reception, but also must have radiation characteristics that discriminate against unwanted signals and minimize interference into other satellite or terrestrial systems. A further function of the antenna is to provide the means of polarization discrimination of unwanted signals. The individual communication system operational parameters dictate to the antenna designer the necessary electromagnetic, structural and environmental specifications necessary for the antenna.

Antenna requirements can be grouped into several major categories: electrical or RF, control systems,

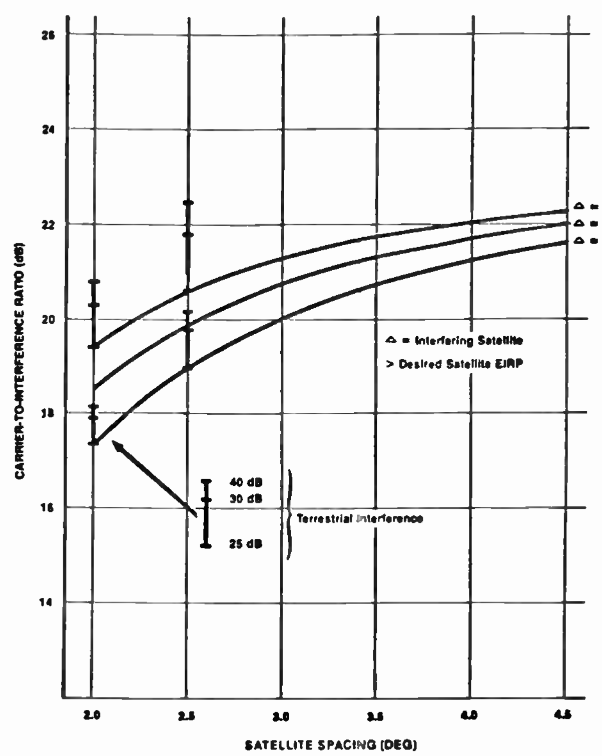


Figure 6.11-34. Carrier-to-interference ratio vs. satellite spacing for a 10-meter uplink and a 7-meter downlink. FX = 6.5 dB, PX = 10 dB. Co-Pol 29-25 Log θ envelope and adjacent satellite polarization interleaving.

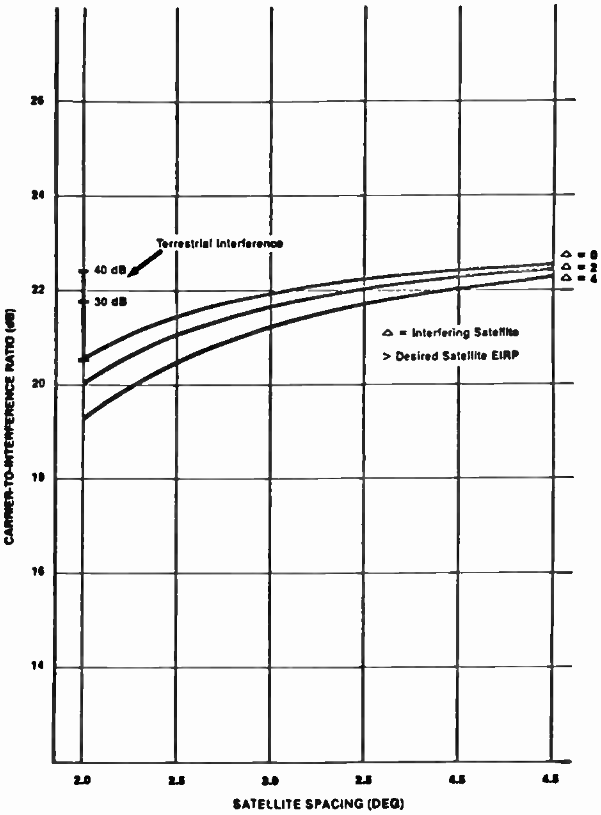


Figure 6.11-35. Carrier-to-interference ratio vs. satellite spacing for a 10-meter uplink and a 10-meter downlink. FX = 6.5 dB, PX = 10 dB. Co-Pol 29-25 Log θ envelope and adjacent satellite polarization interleaving.

pointing and tracking accuracies, environmental and miscellaneous requirements such as radiation hazard, primary power distribution, etc. Table 6.11-5 summarizes many of the more important parameters of an earth station antenna.

Electrical Performance

The primary electrical specifications of an earth station antenna are gain, noise temperature, VSWR, power rating, receive/transmit group delay, radiation pattern, polarization, axial ratio, isolation and *G/T*. All of the parameters except the radiation pattern are determined by the system requirements. The radiation pattern should meet the minimum requirements set by the FCC and/or the ITU-R. Earth stations that operate in a regulated environment in the United States domestic system must meet the requirements set forth in the FCC regulations for earth station antennas pertaining to antenna aperture diameter, sidelobes, and/or radiated power density (see Part 25, paragraph 25.209 of the FCC Rules and Regulations).

The desired radiation properties to satisfy the communication system design dictate the choice of the type of antenna to be employed as an earth station. The three most important radiation properties are gain, sidelobe performance and noise temperature. Most earth station antennas are designed to maximize gain and minimize noise, thereby maximizing *G/T*. These two criteria have led to the predominance of reflector type antennas for earth station applications although other types of antennas such as arrays have been used.

Types of Earth Station Antennas

Several types of earth station antennas are now in use within the United States and abroad. These antennas can be grouped into two broad categories: *single beam antennas and multiple beam antennas.* A single beam earth station antenna is defined as an antenna which generates a single beam which is pointed toward a satellite by means of a positioning system. A multiple beam earth station antenna is defined as an antennas that generates multiple beams by employing a common reflector aperture with multiple feeds illuminating that aperture. The axes of the beams are determined by the location of the feeds. The individual beam identified with a feed is pointed toward a satellite by positioning the feed without moving the reflector. The dual-frequency antennas may be considered another class of antennas as they produce two coincident simultaneous beams and as such are categorized as single beam antennas.

Single Beam Antennas

The majority of the earth station antennas in use today are single beam antennas. Single beam antenna types used as earth stations are paraboloidal reflectors with focal point feeds (prime focus antenna), dual

Table 6.11-5
General Considerations for Earth Station Antenna Design.

Electrical Performance	Mechanical Performance	System Considerations
Frequency (Bandwidth)	Angular Travel	Operational Function
Gain	Drive Speed and Acceleration	Local and/or Remote Operation
Noise Temperature	Pointing and Tracking Accuracies	Availability and Maintainability
Radiation Pattern	Compatibility and Environmental Conditions	Design Lifetime
Polarization	Reflector Surface Accuracy	Interface Conditions with Other Subsystems
Axial Ratio	Physical Dimensions	
VSWR	Weight	
Power handling Capability		
Port-to-port isolation		
Out-of-band Emissions		

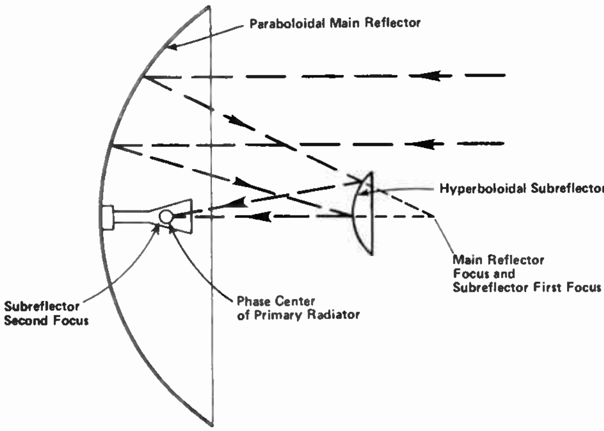


Figure 6.11-36a. Cassegrain antenna geometry.

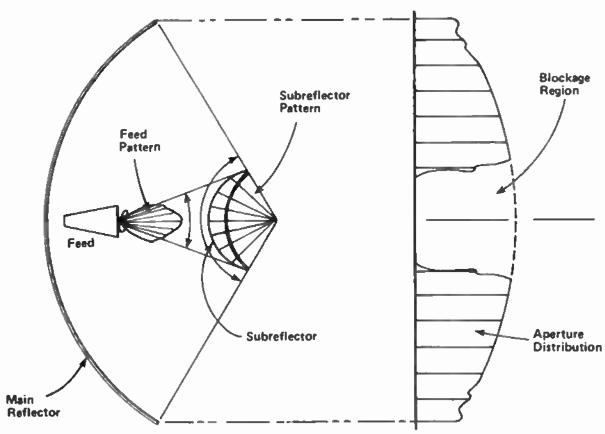


Figure 6.11-36b. Aperture distribution of a cassegrain antenna.

reflector antennas such as the Cassegrain and Gregorian configurations, horn reflector antennas, offset-fed paraboloidal antennas and offset-fed, multiple reflector antennas. Each of these antenna types has its own unique characteristics and the advantages and disadvantages have to be considered when choosing them for a particular application.

Axisymmetric Dual Reflector Antennas

The predominant choice of most system operators has been the dual-reflector Cassegrain antenna. Cassegrain antennas can be divided into three primary types:

- The classical Cassegrain geometry employing a paraboloidal contour for the main reflector and a hyperboloidal contour for the subreflector (Figure 6.11-36). The paraboloidal reflector is a point focus device with a diameter D_p and a focal length f_p. The hyperboloidal subreflector has two foci. For proper operation, one of the two foci is the real focal point of the system and is located coincident with the phase center of the feed; the other focus, the virtual focal point, is located coincident with the focal point of the main reflector.
- A geometry consisting of a paraboloidal main reflector and special shaped, quasi-hyperboloidal subreflector. The geometry of Figure 6.11-37 is appropriate for describing this antenna. The main difference between the classical Cassegrain and this antenna is the subreflector has been designed such that the overall efficiency of the antenna has been enhanced, thereby yielding improved gain performance. This technique is especially useful with antenna diameters of approximately 30–100 wavelengths; for example, a 5 m antenna in the 6/4 GHz frequency band.
- A generalization of the Cassegrain geometry consisting of a special-shaped, quasi-paraboloidal main reflector and a shaped, quasi-hyperboloidal sub-

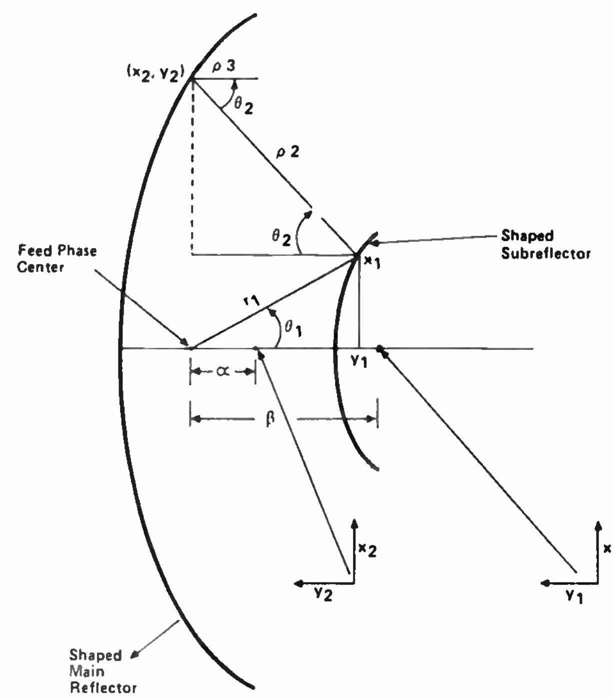


Figure 6.11-37a. Dual-shaped reflector geometry.

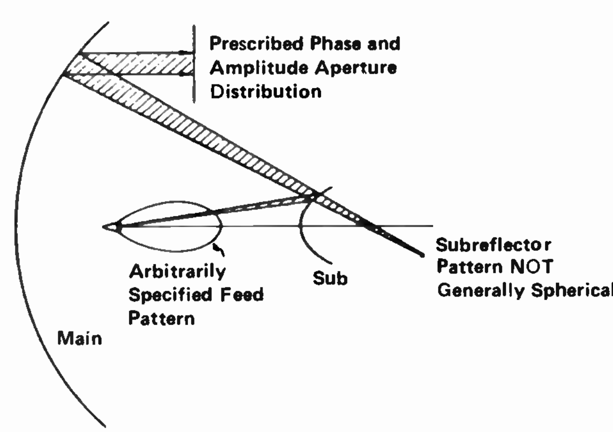


Figure 6.11-37b. Circularly symmetric dual-shaped reflectors.

reflector. The subreflector is shaped to redistribute its incident energy such that the illumination of the main reflector is optimized for high gain and desired radiation pattern. The main reflector is then shaped to correct the phase of the aperture field such that it is in-phase. The feed must have a high beam efficiency and its radiation pattern should be circular symmetric. This technique allows the antenna designer to synthesize the surfaces to achieve an arbitrary aperture distribution.

The dual reflector antenna offers excellent gain performance and for aperture sizes larger than approximately 75 wavelengths the sidelobe performance can meet the FCC pattern requirements. Dual reflector designs are employed for earth station antennas for apertures as small as 50 wavelengths to as large as 500 wavelengths.

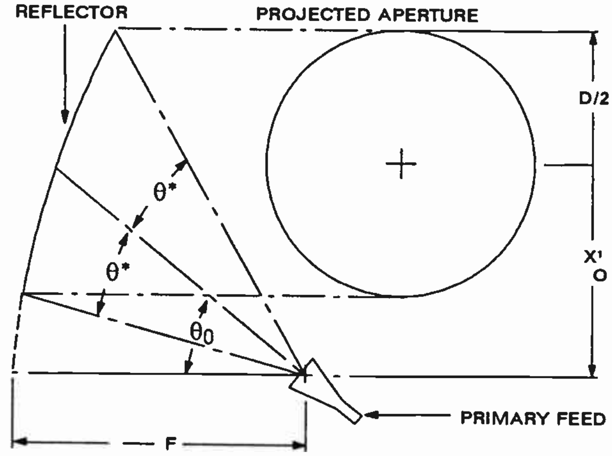


Figure 6.11-38. Single offset reflector antenna.

Prime Focus Fed Paraboloidal Antennas

The prime focus fed paraboloidal (PFFP) antenna is another of the most often employed antennas for earth stations. This type of antenna can have excellent sidelobe performance in all angular regions except the spillover region around the edge of the reflector, but even in this region the pattern requirements of the FCC can be met. This antenna configuration has a lower cost than dual reflector antennas and offers a good compromise choice between gain and sidelobes. Its basic limitations are in its location of the feed for transmit applications and for aperture sizes less than approximately 30 wavelengths, the blockage of the feed and the feed support structure raises the sidelobes with respect to the main beam such that it becomes exceedingly difficult to meet the FCC sidelobe requirements. The PFFP antenna is used for many receive only earth station antennas as well as for transmit/receive applications when only one transmit polarization is required.

Offset-Fed Reflector Antennas

The offset-fed reflector antenna, Figure 6.11-38, has been used primarily in small aperture antennas for VSAT applications. The offset-fed reflector antenna can employ a single main reflector or multiple reflectors, with two reflectors the more prevalent of the multiple reflector designs. The offset, front-fed reflector, consisting of a section of a paraboloidal surface, eliminates the direct aperture blockage from the feed and feed supports and minimizes diffraction scattering by removing the feed and feed support structure from direct illumination of the aperture current distribution. The sidelobe patterns of small apertures can meet the FCC and ITU-R requirements. The limitations of the offset-fed single reflector antenna are in its polarization performance, reduced cross-polarization performance off-axis for linear polarizations and beam squints in opposite directions for two orthogonal circular polarizations. The offset geometry typically means higher manufacturing cost except in the case of small, single piece reflectors such as employed in the VSAT applications.

The offset dual reflector antenna, Figure 6.11-39, can be designed to have all the desirable characteristics of an axisymmetric antenna with increased gain and lower sidelobes. The polarization problems associated with the single offset reflector design can also be compensated for with a two reflector antenna design. The only disadvantages of the offset dual reflector antenna are its high cost of manufacturing for large apertures consisting of multiple sections and the complexity of its mount geometry and its associated high cost.

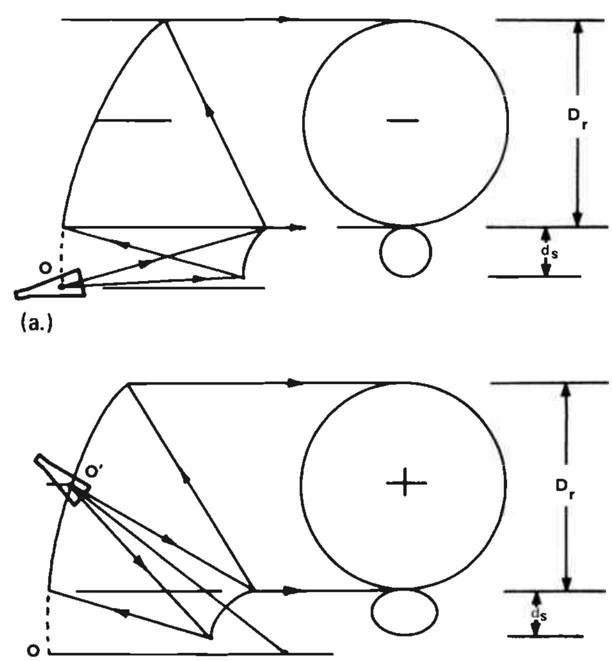


Figure 6.11-39. Offset dual-reflector geometries a) double-offset geometry, (feed phase center and paraboloidal vertex at 0); b) open cassegrainian geometry (feed phase center located at 0'; paraboloidal vertex, at 0).

Beam Waveguide Antennas

The beam waveguide antenna, Figure 6.11-40, utilizes a beam waveguide transmission system to minimize the loss between a multiple reflector antenna and its feed which is located below the rotating axis of the positioning system. This configuration allows the HPA and LNA subsystems to be housed indoors and convenient for maintenance and repair. The beam waveguide design is applicable for very large apertures and is very costly to manufacture. The use of beam waveguide systems has been primarily for 30 m C-band (*Intelsat*) Standard "A" earth stations and for some TT&C earth stations.

Multiple Beam Antenna

Several multiple beam antenna (MBA) configurations, are used for earth station applications (see Figure 6.11-41). These include the spherical reflector, the torus antenna and a class of offset-fed, Cassegrain antennas. All of these configurations employ multiple feeds to generate the multiple beams. The multiple feeds must be physically small such that the individual beams may be pointed at desired satellites. When the desired satellites are spaced as close as 2° apart, the MBA antenna may not be practical. The obvious advantage of the MBA antenna is that a single antenna installation can transmit or receive signals to several satellites simultaneously. The disadvantages are the complexity of the feed arrangements for maintaining pointing for several satellites at the same time when the primary antenna aperture, the main reflector, remains fixed with respect to the earth's coordinates and the stringent requirements for the initial installation of the antenna system. As a result of these disadvantages the MBA antenna is not often employed as an earth station antenna.

Reflector Feed Configurations

There are many different feed configurations used in earth station antennas. The feed configurations are typically classified by the number of transmit and receive ports available. The frequency bands of operation are those specified above for FSS operation in the United States or the appropriate FSS bands for international services. Note that C-band or Ku-band refers to the frequency segments for both transmit and receive bands.

The more popular feed systems are classified as follows:

Two-Port Feed configuration may have two orthogonally polarized receive ports or a single transmit port and a single receive port. The transmit and receive ports may be either co-polarized or cross-polarized with respect to each other. The two-port feed are available in either C-band or Ku-band.

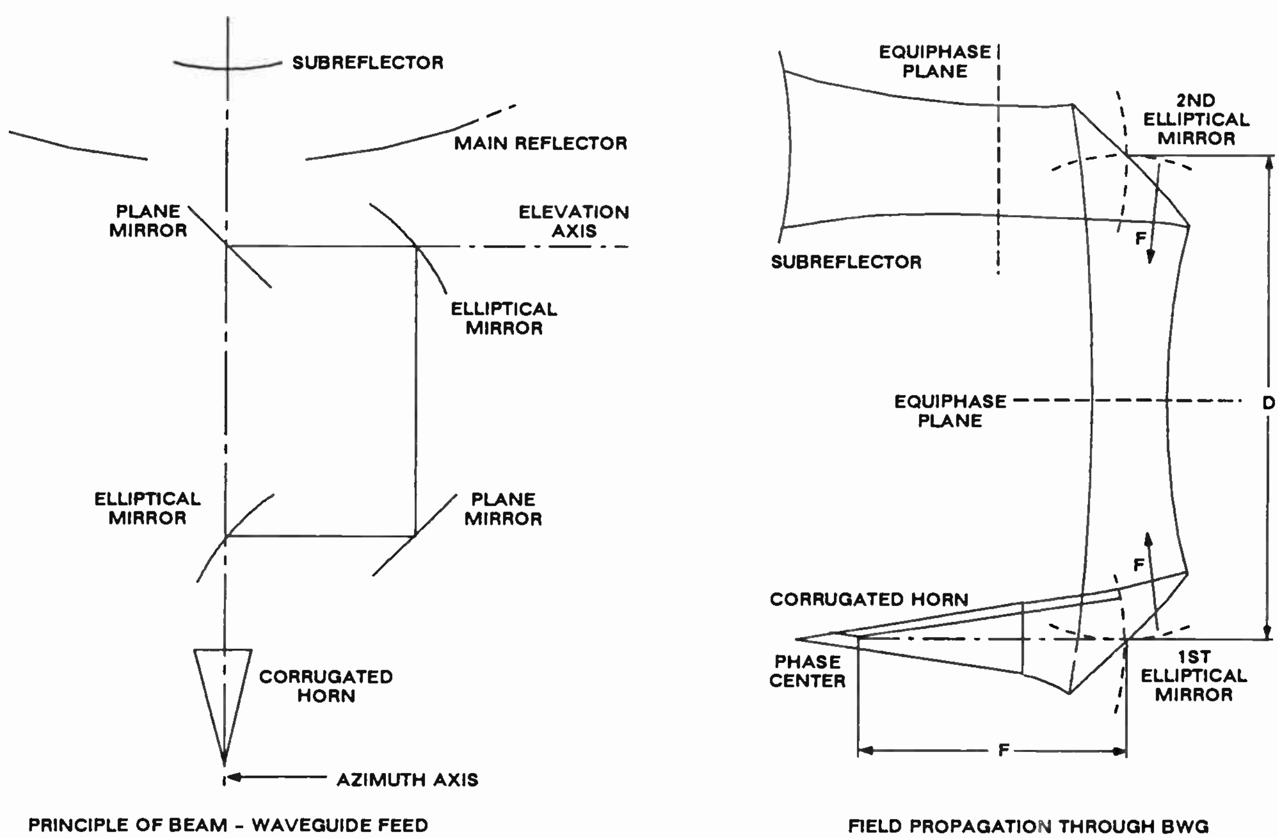
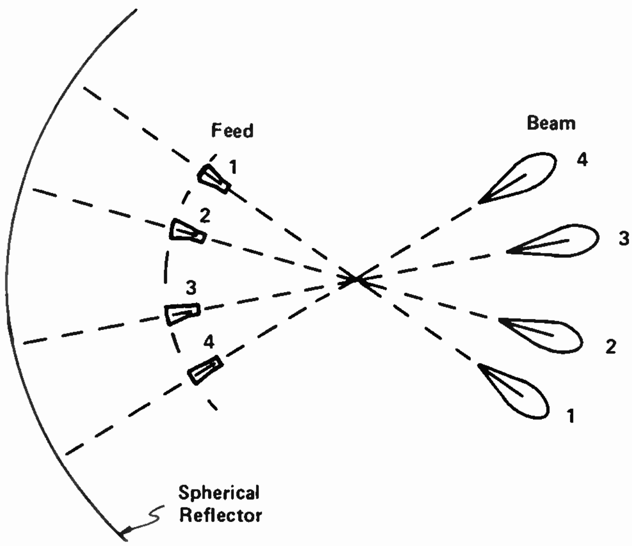
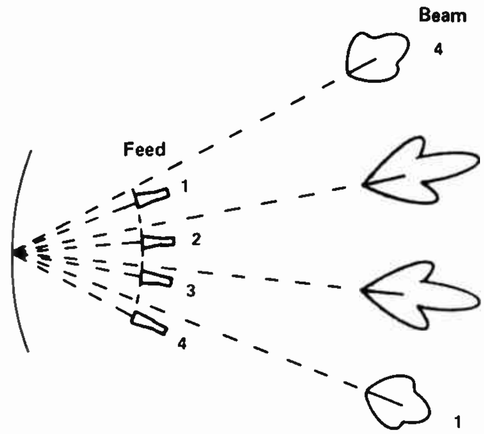


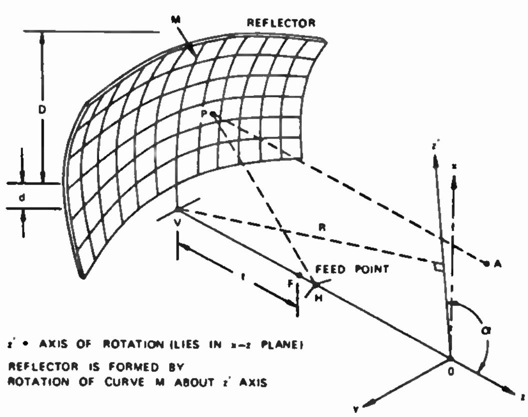
Figure 6.11-40. Beam waveguide antenna.



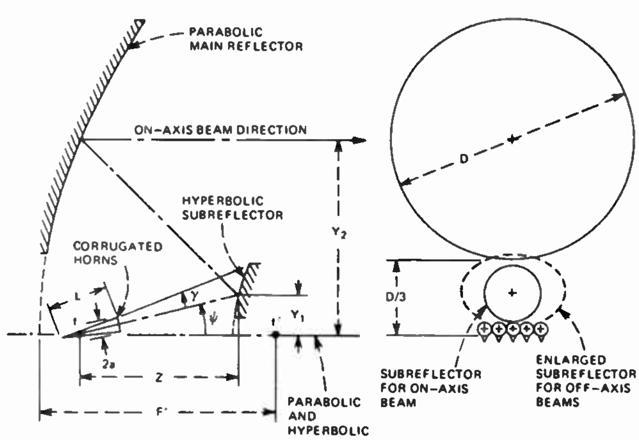
Conventional spherical multibeam antenna
using extended reflector and multiple feeds



Alternative spherical multibeam antenna using
minimum reflector aperture with scanned beam feeds



Torus–antenna geometry
(Copyright 1974, *COMSAT Technical Review.*
Reprinted by permission.)



Geometry of the offset–fed multibeam Cassegrain antenna.
(Copyright, 1974, *American Telephone and Telegraph Company.*
Reprinted by permission.)

Figure 6.11-41. Multiple beam configuration.

Three-Port Feed has two receive ports and a single transmit port. The receive ports provide for two orthogonal polarizations. The three-port feeds are available in either C-band or Ku-band.

Four-Port Feed provides for dual polarization capability for both the transmit and receive applications. This feed configuration is also referred to as a frequency reuse feed. The four-port feed configurations are available for either C-band or Ku-band.

Dual Band Feed provides for the simultaneous reception of C-band signals and Ku-band signals from a hybrid satellite. The dual band feeds are typically a single aperture, that is the C-band and Ku-band radiating apertures occupy the same space and usually sacrifice gain and sidelobe performance to provide the dual frequency operation. This is true for the single reflector and dual reflector designs.

An alternate configuration that does not sacrifice radiation performance utilizes a dual reflector geometry where the subreflector is a frequency selective surface (fss). This configuration typically uses a prime focus C-band feed, a fss subreflector which is transparent to C-band and reflective for Ku-band and a Ku-band dual-reflector feed. The Ku-band feed may be as simple as a single-port feed to a full frequency reuse, four-port feed, whereas C-band is a prime focus, receive-only feed.

Mechanical Performance

The mechanical design of an earth station antenna must provide the structural integrity to accurately point the antenna beam towards the desired satellite and to maintain the pointing accuracy within the environmental conditions for the locale. Further, the mechanical design of the antenna must ensure the required tolerance of the radiating surface such that the radiation performance of the antenna is not compromised. The antenna pedestal must also provide the means to steer the antenna beam to the satellites of interest.

The location and size of an earth station antenna system (antenna, pedestal or mount, electronics and control housing) usually make it subject to local building codes. The code which is almost universally accepted is the *American Standard Building Code Requirements for Minimum Design Loads in Buildings and Other Structures, ANSI A58.1*. Paragraph 1.3 of that Standard states:

> Buildings or other structures and all parts thereof, shall be designed and constructed to support safely all loads, including dead loads, without exceeding the allowable stresses (or ultimate strengths when appropriate load factors are applied) for the materials of construction in the structural members and connections. When both wind and earthquake loads are present, only that one which produces the greater stresses need be considered and both need not be assumed to act simultaneously.

The loads that must be safely supported by an earth station antenna system are the weight of the antenna and the attached equipment, the expected ice and snow load, earthquake load and the wind load. Of these, the wind load is usually the largest single contributor to the stress and deflection of the structure.

Earth station antennas have a specification that is variously called maximum wind, survival wind, or withstand wind. These terms should be considered synonymous. At the manufacturer's specified survival wind the system must be safely supported without exceeding the allowable stresses for the materials. Survival wind, as defined herein, when combined with ice and dead weight results in the *Design Load* as defined in Standard EIA-222C. In addition to survival wind, two other sets of wind conditions are usually specified, the operational wind velocity and the drive-to-stow wind velocity. The operational wind velocity is the maximum value at which the antenna system fully meets the performance specifications. The drive-to-stow wind velocity is the maximum value the antenna may be driven through the azimuth and elevation actuators to the prescribed stow position (usually zenith).

Positioning Systems

There are two broad classes of positioning systems used for satellite earth station antennas. One class consists of orthogonal two-axis configurations; the other is the one-axis or single-axis configuration. The two-axis systems are characterized by the orientation of the lower most axis with respect to the earth. A two-axis system having its lower axis perpendicular to the ground is called an *elevation-over-azimuth* (see Figure 6.11-42a). One that has its lower axis parallel to the ground (see Figure 42b) is called *X-Y*. One that has its lower axis parallel to the earth's axis of rotation is called an *hour angle-declination or polar* (see Figure 6.11-42c). Each of the three positioning systems has the beam-axis or pointing direction perpendicular to the upper axis. Providing there are no physical limitations, all three types can theoretically point in any direction.

Development of single-axis antenna mounts was brought about by efforts to reduce the costs and to simplify the positioning of the antenna beam with respect to the geostationary orbit. This geometry has restricted applications due to its lack of capability to follow satellites in inclined orbits and its inherent error as the pointing transverses the geostationary arc.

The elevation-over-azimuth positioner has become the choice for most systems. Figure 6.11-43 is a graph of azimuth and elevation angles versus a particular site latitude and longitudinal difference between the satellite and the site. The horizontal and vertical rectangular coordinates are site latitude and difference longitude, respectively. The curved lines running toward and labeled at the top of the graph are the required azimuth angles (add 180° if satellite is west of site, subtract from 180° if satellite is east of site). The curved lines running toward and labeled at the left margin (down to 15°) are the required elevation angles. (The elevation lines of 10° and below are labeled at the top of the graph.) To determine the required azimuth and elevation pointing angles, find the satellite site longitudinal difference and move vertically on this line until it intersects the horizontal latitude line. At this intersection interpolate between bounding azimuth and elevation curves for the requiring angles.

The azimuth and elevation angles to a particular geostationary satellite can be calculated using the satellite longitude, Z; the site longitude, Y; the site latitude X; and the following equations:

$$C = Z - Y \quad \text{degrees} \qquad [34]$$

$$A(\text{azimuth}) = 180 + \tan^{-1}[\tan(C)/\sin(X)] \quad \text{degrees} \qquad [35]$$

$$E(\text{elevation}) = \tan^{-1}\left\{\frac{\cos(C)*\cos(X) - 0.15126}{[\sin^2(C) + \cos^2(C)*\sin^2(X)]^{1/2}}\right\} \quad \text{degrees} \qquad [36]$$

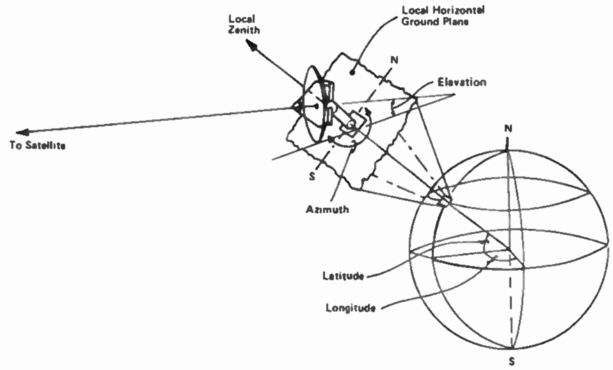


Figure 6.11-42a. Elevation-over-azimuth geometry.

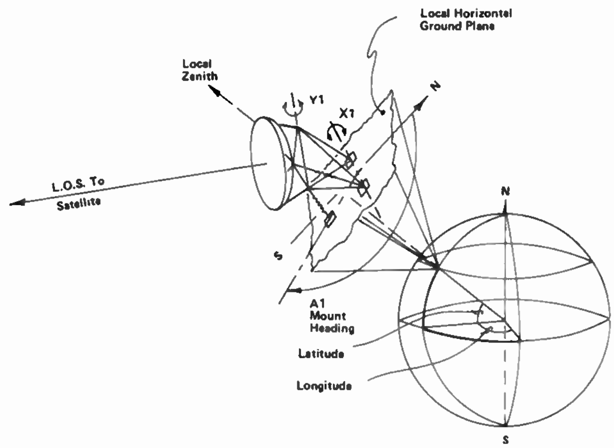


Figure 6.11-42b. X-Y mount geometry.

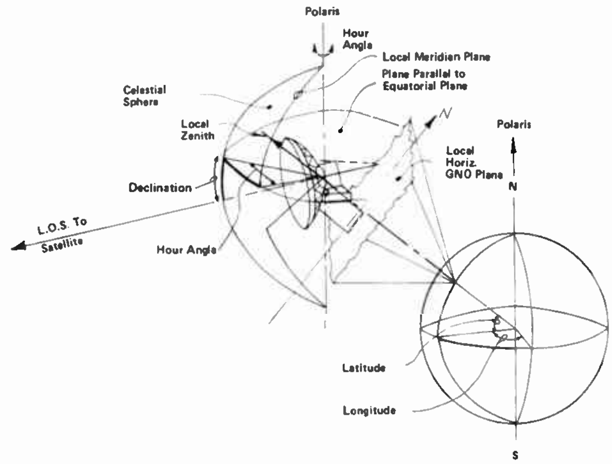


Figure 6.11-42c. Polar mount geometry.

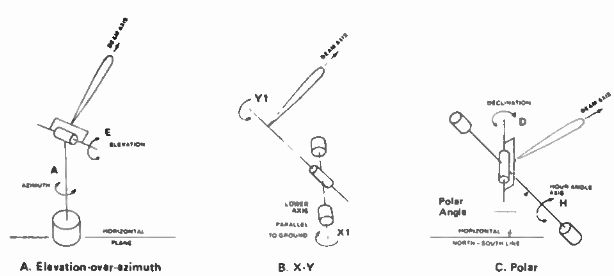


Figure 6.11-42d. Schematic illustration of two-axis earth station positioning systems.

All of the other environmental conditions at a particular site should be addressed, such as effects of solar radiation, lightning strikes, damage by salt water, acid rain and pollution gases, in the planning and implementation of the earth station system.

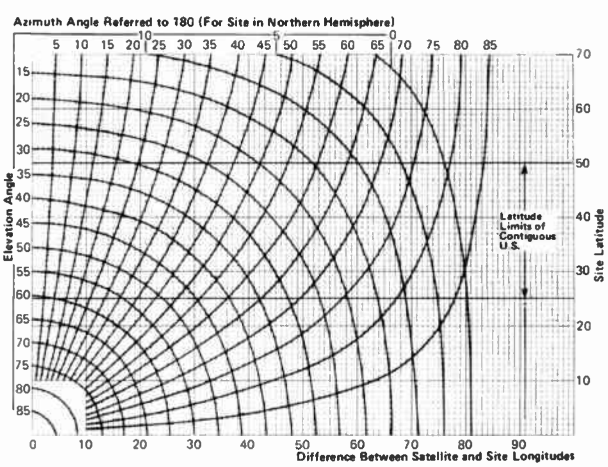


Figure 6.11-43. Universal azimuth-elevation look angles.

Pointing and Tracking

The pointing and tracking accuracy are two very important considerations for a earth station antenna system. Pointing accuracy is defined as the precision with which an antenna can be held (for fixed position antenna) or steered under the specified operating conditions. The pointing error is a measure of pointing accuracy and is defined as the space angle difference between the command vector and the actual position of the antenna communication RF axis. Pointing error is usually specified to less than 0.2 of the half power beamwidth (HPBW), of the antenna in the transmit frequency band. Tracking accuracy is the precision with which an antenna can track a source under specified operating conditions. The tracking error is a measure of tracking accuracy and is defined as the space angle difference between the communication RF axis of the antenna and the vector to the RF source. Tracking error is usually specified to be less than 0.1 of the HPBW. Table 6.11-6 lists the sources of error that should be considered in an overall budget or calculation of pointing accuracy and tracking accuracy.

Many earth station systems operate in the point mode, that is, there is no requirement for automatic tracking of the satellite. This condition exists when

Table 6.11-6
Pointing and Tracking Error Budget Terms.

Pointing Error Budget	
Velocity Lag	Breakaway friction
Wind up of gear train	Secant potentiometer
Angle encoder	Tachometer
Angle encoder coupling	Amplifier drift
Level	Amplifier bias
North alignment, initial zeroing of encoders	Motor cogging
AZ-EL axis orthogonality	Backlash
RF-EL axis orthogonality	Servo dead zone
Reflector alignment	Servo noise
Structural Distortion	Axis wobble
Gravity	Radome diffraction
Ice	Boresight shift vs. polarization
Wind	Boresight shift vs. frequency
Thermal	
Acceleration	
Foundation displacement	
Acceleration lag	

Tracking Error Budget	
Velocity lag	Wind torque (servo)
Tracking receiver	Null axis—beam axis alignment
Acceleration lag	Breakaway friction
Tachometer	Motor cogging
Amplifier drift	Backlash
Servo noise	

the satellite orbital location is maintained within a small fraction of a degree (<0.1°) and when the earth station antenna half power beamwidth is sufficiently broad (>0.5°). Automatic tracking may become necessary as the antenna becomes large in terms of wavelengths (very narrow RF beam) or if the satellite is allowed to transverse an inclined orbit. The complexity of the tracking system is determined by the overall system accuracy requirements and the allowance in EIRP and *G/T* that is budgeted for impaired operation.

A hierarchy of pointing and tracking systems is as follows:

a) Initial fixed pointing is satisfactory (receive only)
b) Repointing of the antenna is required to switch between various satellites or to correct for satellite motion
c) Tracking is required to correct for satellite drift. Satellite position vs. time is known and program track is satisfactory
d) Automatic tracking is necessary but can be satisfied by a simple step-track system
e) Full automatic tracking is necessary (extended inclined orbits)

The simple step-track system is satisfactory for most satellite communication applications when automatic tracking is required. The step-track systems generate tracking information by moving the RF beam in several steps, comparing the signal level, deciding the proper direction to move for the next step and then continuing this process until the RF signal is maximized. The step-track system uses a very low frequency servo loop and therefore will not track out such disturbances as wind. Step-track can be susceptible to fade conditions unless the sampling circuitry is preset to cut off when a large signal loss is evident. Steptrack may also be augmented with a program track mode whereby the satellite movement is memorized and then followed by a memory command circuit.

Fully automatic tracking systems are typically used for TT&C earth stations or for those earth stations operating under extreme conditions with very narrow RF beamwidths. The automatic tracking configurations include conical scan, electronic beam scanning (Single Channel Monopulse, ESCAN) and three channel monopulse. The electronic scanning and three channel monopulse techniques offer the advantage of providing a data channel and a transmit channel without tracking modulation superimposed on the signals. This is not possible with the conical scanning technique.

Transmit Electronics

The transmit subsystem consists of equipment from baseband to the high power RF amplifier. These electronics include baseband processors for combined video and audio signals, modulators, upconverters and high power RF amplifiers. Subcarrier modulators are used to insert various audio signals on the video signal prior to the wideband modulation of the signal.

Analog Video Services

Baseband Processor

The incoming analog video signal is first processed by the baseband processing module. The video signal is preemphasized for either 525-line (NTSC) or 625-line (PAL/SECAM) operation and passed through a low pass roofing filter. Preemphasis acts to improve the output video signal-to-noise ratio by compensating for the increase in noise density with frequency (triangular noise) which is characteristic of the receiver's FM demodulator. (The preemphasis is removed by a deemphasis network after the receiver discriminator).

Energy dispersal modulation is applied to the incoming video signal. Satellite transmissions of video signals are processed in this manner to disperse the RF spectrum thus preventing concentration of energy. This reduces interference with terrestrial microwave and other satellite links and reduces intermodulation among the multiple carriers which exist in a real satellite. Energy dispersal modulation is applied using a triangular waveform with apexes located at the vertical intervals of the video signal.

Subcarrier Modulator

Each subcarrier modulator inputs an audio signal and modulates that signal onto a carrier positioned in the range between 5.0 and 8.5 MHz. Typical subcarrier frequencies are 6.2 MHz and 6.8 MHz. Most video exciters synthesize the center-frequency of the subcarrier with a resolution (stepsize) of 10 kHz. Generally, a subcarrier modulator will allow for selection of preemphasis. Frequency deviation (the amount of frequency modulation applied) is also typically adjustable between 50 kHz and 500 kHz peak.

Wideband Demodulator

The wideband modulator frequency modulates the composite signal (baseband video and subcarriers) onto a 70 MHz IF carrier. The frequency deviation is adjustable up to 15 MHz peak. The modulation bandwidth of the output signal for NTSC video is approximately 36 MHz for full transponder application.

Upconverter

The upconverter converts the modulated 70 MHz IF signal to a frequency range compatible for satellite transmission. Generally, the signal is converted to C-band or Ku-band.

Video Exciters

Most commercial-use video signals transmitted by satellite are of two types: a broadcast quality commercial television signal or compressed, digitally encoded representation of a video signal.

Compressed digital video is typically used by private business television networks for training and internal communication on secure channels.

Analog Video Exciter

A system that transmits conventional analog video, as used for broadcast quality transmission, will use a video exciter. The video exciter inputs a set of baseband signals and outputs a single signal suitable for transmission over a satellite link. The center-frequency of the output signal generated by the video exciter is a frequency in the range of the frequency band in use (5.925 to 6.425 GHz for C-band, 14.0 to 14.5 GHz for Ku-band). The set of baseband signals typically consists of a video signal and one or more audio signals. In some cases, one of the audio signals is used to encode a data control channel. A block diagram of a typical video exciter is shown in Figure 6.11-44.

Digital Video Codecs

A system used for transmission of compressed digital video uses a video codec (coder/decoder) to sample the applied analog video and audio waveforms, generating a digital representation of the input signals. Additionally, most codecs perform compression to reduce the bit rate of the signal. At least two formats of digital video are represented as broadcast quality. First, the D2 format encodes video at approximately 112 Mbps. This format is sometimes used for studio tape playback and recording of programming. Second, a 45 Mbps system is available for transmission of video over DS3 digital circuits. DS3 is a standard rate (45.3 Mbps) available from common carriers.

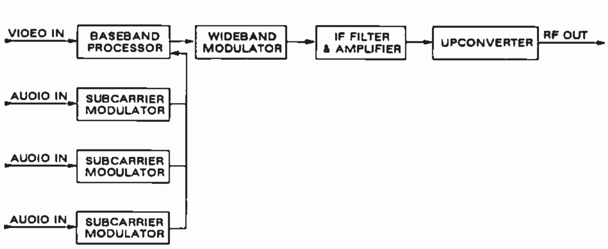


Figure 6.11-44. Typical video exciter block diagram.

High-Power Amplifier

The high power amplifier (HPA) amplifies the RF output signal from the upconverter to the required power level for transmission to the satellite. Amplifiers for satellite video applications are typically sized in the range from 1 W to 3 kW. Amplifiers in the 1 to 10 W range are available in solid-state configurations. Travelling wave tube (TWT) amplifiers are available in configurations up to approximately 750 W. For power levels above 750 W, Klystron tube amplifiers are used.

Video uplink systems using U.S. domestic C-band satellites usually employ 3.0 kW Klyston HPA's. However, newer applications using higher power Ku-band satellites typically use lower power TWT amplifiers. Three hundred watt TWT amplifiers are commonly used in occasional video service applications such as education, private business networks and satellite news gathering.

The HPA usually contains bandpass filters to reject harmonics and power sampling circuits for monitoring the output transmit power and the reflected power from the antenna. Often, protection circuitry is added to turn off the HPA when the reflected power exceeds some predetermined level.

Automatic Transmit Identification System

With the increase in uplink activity in the 1980's, there has been a growing problem with interference between carriers. Most incidents are unintentional, but there have been reported cases of malicious interference. To counter the problem, the FCC introduced a requirement that video uplinks must incorporate an automatic transmit identification system (ATIS), to identify the source of the transmission. The ATIS signal is a FM subcarrier positioned at 7.1 MHz. The ATIS subcarrier contains a message composed of international Morse code characters to identify the source of the signal and provide a telephone number for communication with its operator. The message is repeated every 30 seconds and includes a unique 10 digit ID code that is unchangeable by the operator. The subcarrier frequency of 7.1 MHz was chosen since it is very close to the second harmonic of the color subcarrier and therefore not useable for any possible revenue source. The subcarrier injection level of −26 dB referenced to the unmodulated main carrier represents a reasonable compromise between ATIS system sensitivity, resistance to interference and power taken from the main carrier. This injection level is approximately 0.05 of the normal level of a monaural TV associated audio subcarrier.

Audio Service

Audio signals are transmitted by satellite in both analog and digital form. Most of the domestic U.S. nationally distributed audio material is delivered in digital format. The satellite distribution encodes program material in digital form at the source and distributes the information in that form.

Other networks (state and regional networks) transmit audio using single channel per carrier (SCPC) systems. These systems frequency modulate the applied audio signal onto a carrier and upconvert the signal to a satellite compatible frequency range.

Digital Audio Transmission

The digital audio system supports four types of signals: voice-grade, 7.5 kHz audio, 15 kHz audio and data. The 7.5 kHz and 15 kHz audio signals are sampled at 16 kHz and 32 kHz respectively. The signal is digitized with a 15 bit converter and u-law companded to 11 bits plus a parity bit. The voice-grade signal is digitized with a continuously variable slope delta modulation (CVSD) scheme producing a digital bit rate of 32 kbps. The resultant set of digital signals are multiplexed into a single T1 (1.544 Mbps) bit streams for transmission to the earth station. The T1 signals are routed to the earth station over redundant paths. The earth station receives the multiple, redundant T1 bit streams and demultiplexes the data into the original set of digital signal components.

These component digital signals are then multiplexed into a 7.68 MHz aggregate bit stream. The 7.68 Mbps data is forward error correction encoded, modulated onto a 70 MHz IF carrier and upconverted/amplified to a C-band or Ku-band frequency for satellite distribution. For more information on the digital transmission of the digital information, see the previous section on digital transmission.

Typical bit rates for a 15 kHz audio channel using conventional pulse code modulation (PCM) is 512 kbps. Companding techniques reduce this bit rate to 384 kbps. The availability of low cost, high performance digital signal processing (DSP) integrated circuits have now resulted in more effective compression techniques for digital audio encoding. Techniques such as sub-band coding and frequency domain transforms are used to provide typical bit rates of 128 kbps for the same 15 kHz audio channel. Thus, DSP techniques have provided dramatic increases in spectrum efficiency.

Receive Electronics

The receive electronics are similar in scope to the above transmit subsystem but, of course, operate in the reverse order. First the incoming RF signal is filtered, amplified, downconverted (optional, depending on the frequency band of operation) and passed to the receiver where the signal is further downconverted, amplified and demodulated to baseband.

Low-Noise Amplifiers/Low-Noise Converters

The first active signal processing of a downlinked satellite signal occurs at the low noise amplifier (LNA) or low noise block/converter (LNB). Traditional C-band broadcast applications have used a LNA mounted at the antenna connected to the indoor electronics

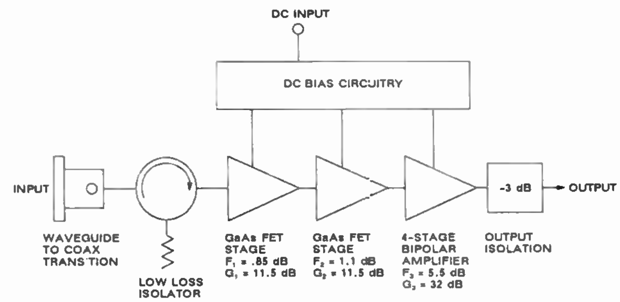


Figure 6.11-45. LNA block diagram.

through a length of coaxial cable. Typical Ku-band systems use a LNB at the antenna that amplifies and downconverts the signal to L-band (950 MHz to 1450 MHz). The functions of the LNA are as follows:

- Provide high gain and low noise to establish high system *G/T*
- Provide transition from antenna waveguide to TEM coaxial cable. Since long waveguide runs are expensive, the LNA is designed to accept a waveguide input and provide a coaxial line output
- Provide adequate mechanical strength to permit mounting directly to the antenna waveguide and to allow connection of a long coaxial cable to the unit
- Provide RFI/EMI tight weatherproof housing for circuitry

A block diagram of a typical LNA is shown in Figure 6.11-45.

Video Receivers

The video receiver takes the received satellite signal and produces a collection of baseband signals. The baseband signals are a video signal and one or more audio signals. Traditionally, broadcast quality applications used C-band video receivers. However, with the increased use of Ku-band and decreasing costs for LNB's, many users have converted to L-band systems. A block diagram of a typical video receiver is shown in Figure 6.11-46. The downconverter converts the input RF signal to an intermediate frequency, typically 70 MHz. The demodulator acquires the modulated video signal from the 70 MHz signal.

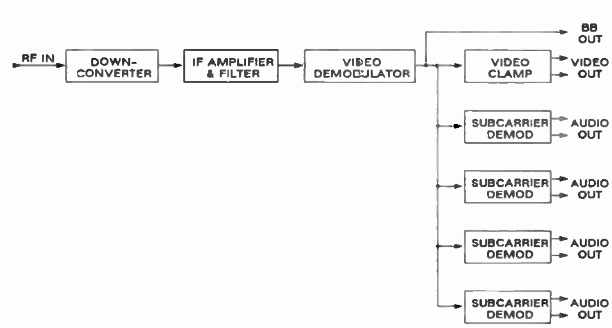


Figure 6.11-46. Typical video receiver block diagram.

Downconverter

The downconverter converts the RF-input signal to a intermediate frequency (IF) prior to demodulation. This intermediate frequency is typically 70 MHz. The downconverter will provide either a single input or multiple inputs for multiple antenna, multiple polarization operation. Single input downconverters require an external relay to select the appropriate polarization input. Modern video receivers typically provide up to four inputs to provide convenient use for two antenna, dual-polarization operation.

The downconverter inputs a signal in range of 3700 MHz to 4200 MHz for C-band operation or in the range of 950 MHz to 1450 MHz (L-band) for C- and Ku-band operation. For the latter case, the LNB downconverts the input RF signal to L-band frequencies.

IF-Filter/Amplifier

The output of the downconverter is routed to the 70 MHz IF filter/amplifier. The signal is first bandpass filtered. In the early days of satellite video services the video receivers provided a single IF filter bandwidth, however, most video receivers offered today provide multiple IF filter bandwidths to provide convenient half transponder as well as full transponder operation. Some video receivers provide up to six IF filter bandwidths. Six IF filter bandwidths are economical because of the availability of surface acoustic wave (SAW) filters for this application.

Demodulator

The filtered, amplified IF signal is fed to the demodulator for FM demodulation. The output of the demodulator is the baseband video and multiple audio subcarriers.

Video Processing

After demodulation, the video baseband contains the 30 Hz triangular energy dispersal waveform which is removed by a circuit referred to as the clamp. A low pass filter is applied to remove the audio subcarrier from the video baseband signal.

Subcarrier Demodulator

Video receivers generally provide up to four audio subcarrier demodulators and provide baseband audio (generally 600 Ω balanced) outputs.

Digital Demodulator/Downconverter

Digital data is extracted from a intermediate frequency carrier by the demodulator section of a modem. (For detailed information about the operation of the modem, refer to the section on digital transmission). The digital demodulator separates the transmitted data stream from the carrier. The transmitted data stream is then processed by the FEC decoder removing the error correction bits, correcting detected errors and outputting the received data and clock signals.

Protection Switching

With the exception of receive only systems most satellite transmission systems contain redundant subsystems to facilitate very high availability specifications. Satellite teleports advertise availability specifications as high as 99.995%. Protection switching implements automatic subsystem redundancy. The protection switch monitors one or more online subsystems for failure. Upon detecting a failure in the subsystem, the protection switch switches the inputs and outputs from the online unit to the backup unit and configures the backup to the configuration of the failed online unit.

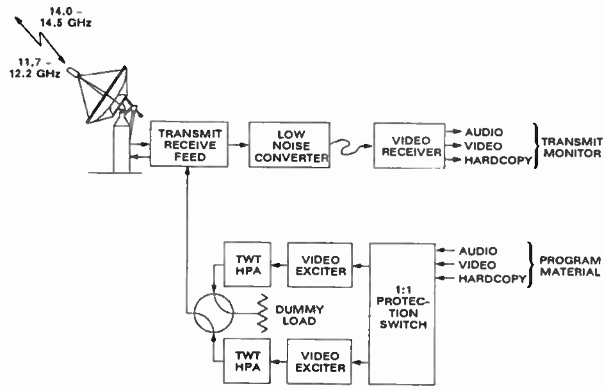


Figure 6.11-47. Redundant Ku uplink protection system.

The configuration of the subsystems protected by a protection switch is referred to as m:n. "m" refers to the number of backups available. "n" refers to the number of online units monitored and protected by the protection switch. "1:1" refers to the simplest configuration—a single backup is available to replace a single online unit. A typical 1:1 protection system for a Ku-band uplink in shown in Figure 6.11-47. A typical 1:1 C-band downlink is shown in Figure 6.11-48. Operation of the 1:1 configuration is simplified in that the backup may be tuned to the same configuration as the online unit; switching to the backup merely requires switching the source of the input and output signals from the online unit to the backup. An example of a larger configuration is "2:6", where two backup units are available to protect six online units. These larger configurations are used for large multiple-channel uplinks. The larger configurations become more cost-effective as the number of online channels increases.

Monitor and Control Systems

Monitor and control refers to systems used to monitor earth station components for failures and provide manual and automatic control of the components.

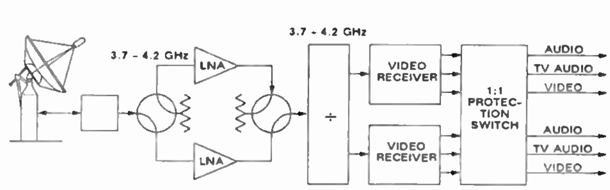


Figure 6.11-48. C-band redundant downlink protection system.

These systems are widely used for a variety of reasons. First, even though most earth station components provide front panel monitor and control functions, there are generally too many earth station components to monitor from the front panels of the respective components. A monitor and control system provides a single point of monitoring and control for the operator thus easing the operator workload allowing the operator to handle more transmissions. Second, many earth stations are located remotely from the studio. A monitor and control system provides remote monitoring and control (operation) of the earth station. Remote operation is facilitated through the use of a low speed data circuit between the earth station equipment and the monitor and control computer in the studio. This circuit is typically a subcarrier on a microwave-radio channel, a dial-up phone circuit with computer modems, or a dedicated EIA-422 hard wire connection for distances less than 1000 m.

Earth Station Control Computer

Monitor and control systems are based upon the use of a general purpose computer executing applications software designed for monitor and control of communications systems. This computer is referred to as the earth station controller.

Operator Interface

Earth station control computers provide a video monitor for display and a keyboard for operator interaction. The system displays earth station status in the form of a hierarchical, graphical display. The upper level display shows a high level block diagram of the earth station. The operator may selected more detailed displays of subsystems by selecting the subsystem symbol from the screen. The earth station components/subsystems are coded with color to indicate state. Typical color conventions for earth stations are as follows:

- Blinking red indicates an unacknowledged alarm
- Red indicates an acknowledged alarm
- Amber indicates ready or standby; Green indicates online/OK/normal.

In some cases, an audible alarm is provided. When a new alarm condition occurs, the audible alarm is turned on. The operator must acknowledge the alarm to turn off the audible alarm.

Earth Station Interface

The computer interfaces to the earth station components through a number of interfaces, including serial, contact closure and other customer supplied interfaces.

- **Serial Interfaces.** The most common interface is a serial ASCII protocol based interface. The interface is usually a asynchronous character oriented scheme utilizing EIA-232C or EIA-422 signal levels. EIA-232C interfaces are used for short cable distances (computer to device) less than 5 m or for connection to a modem. EIA-422 interfaces may be used for cable lengths of up to 1000 m and are used in multidrop mode thus allowing many devices to share a single interface port.
- The protocols are usually ASCII based since there is no satellite communication standard and most vendors use their own version of a interface protocol. Thus, earth station computers must offer a number of serial equipment interface ports and also support a variety of protocols on those ports.
- **Contact Closure Interfaces.** The second type of interface commonly used is the contact closure interface. Older components typically provided contact closure interfaces and offered no serial interface functions. Many of these components are still in use today. In addition, there are components such as waveguide switches and shelter alarms (intrusion, air conditioner, emergency generator, etc.) that only offer a contact closure interface. For example, a waveguide switch may provide two status points and two control points. Thus, earth station computers must provide some method of monitoring and controlling status and contact closure controls. Some vendors offer systems that connect the contact closure directly to the computer. Others offer general purpose interfaces that reside in the earth station and interface to the earth station computer through a serial interface.

Status inputs to the earth station computer are usually optically isolated. The earth station computer supplies the optical isolator. The monitored device sinks current through the isolator to indicate one of two states. The alternate state is no current flow.

Control outputs are of two types. The most flexible interface is the Form C output. The Form C output provides a common connection and a normally opened (NO) and normally-closed (NC) connection. The second type of control output is the open collector output. This interface provides a connection to the collector of a transistor to sink current. One control state is with the transistor on thus sinking current from the controlled device. The alternate state is with the transistor off.

- **Vendor Supplied Computer Interfaces.** A third type of earth station interface is a vendor supplied, device specific contact closure adaptor. Most TWT amplifiers provide contact closure based remote monitor and control interfaces. Additionally, forward and reflected power indications are provided by a signal with voltage level proportional to power. Additionally, some TWT amplifier's require a analog current signal to control the attenuator. However, most TWT amplifier vendors offer a interface adaptor that converts the contact closures to a serial protocol based EIA-232-C or EIA-485 interface. These adaptors reside in the rack with the TWT amplifiers and connect to the earth station computer through a serial connection.

CONCLUDING REMARKS

There are many additional aspects of the design, installation, operation and maintenance of an earth station antenna system that have not been discussed in this chapter. The site selection and preparation are, in particular, critical to the successful operation of the system as well as the foundation design. Details of this aspect of the earth station design should be accomplished with the help of qualified consulting engineers and frequency coordination experts. The operations building and/or equipment houses should be in close proximity to the earth station, if possible, but remote operation is possible. The power requirements for the earth station should also be carefully planned to provide adequate power for the electronics, including the transmitter equipment and any power required for antenna deicing where applicable.

BIBLIOGRAPHY

Cook, J.H. Jr., & Hollis, S., eds., *Communications Symposium '83 Notebook,* Scientific-Atlanta, Inc., Atlanta, GA., November, 1983.

Freeman, R.L., Reference Manual for Telecommunications Engineering, Wiley-Interscience Publications, John Wiley & Sons, Inc., New York, N.Y.

Inglis, Andrew F., ed., *Electronic Communications Handbook,* McGraw-Hill Book Co., New York, NY., Chapters 2, 5, 6, & 17.

Ippolito, Louis J., Kaul, R.D., Walace, R.G., *Propagations Effects Handbook for Satellite Systems Design,* NASA Reference Publication 1082, December 1981.

Johnson, R.C. & Jasik, H., Earth Station Antennas, *Antenna Engineering Handbook,* Cook, J.H., Jr., McGraw-Hill Book Co., New York, NY., Chapter 36.

Jordan, Edward C., Reference Data for Engineers: Radio Electronics, Computers and Communications, Howard W. Sams & Co., Inc., Indianapolis, IN.

Martin, James, *Communications Satellite Systems,* Prentice-Hall, Inc., Englewood Cliffs, NJ.

Morgan,W.L. & Gordon, G.D., *Communications Satellite Handbook,* Wiley-Interscience Publications, John Wiley & Sons, Inc., New York, NY.

Pratt, T. & Bostian, C.W., *Satellite Communications,* John Wiley & Sons, New York, NY.

Pritchard, W.L., & Sciulli, J.A., *Satellite Communication Systems Engineering,* Prentice-Hall, Inc., Englewood Cliffs, NJ.

6.12
LOW POWER TELEVISION

ROBERT M. UNETICH
ITS CORPORATION, MCMURRAY, PA

Additional material contributed by
PETER TANNENWALD
IRWIN, CAMPBELL & TANNENWALD, P.C., WASHINGTON, DC

INTRODUCTION

Conventional full service television broadcasting is built on the concept of high powered single-channel stations with wide area coverage. This service is presently supplemented by two lower powered terrestrial television technologies: low power television/translators operating on conventional TV channels and multichannel multipoint distribution service/instructional television fixed service (MMDS/ITFS) operating in the 2500–2690 MHz band.

Low Power Television and TV Translators

The FCC first authorized TV translators in 1956 as a way to bring television service to remote areas or areas shielded by terrain obstacles. Program origination (was prohibited) until 1982, when low power television (LPTV) service was created as a way to introduce new local programming services to markets that might otherwise not have television service. There are currently some 2,000 licensed LPTV stations and 5,000 licensed translators, several hundred of which are in Alaska. In addition, the FCC authorizes conventional TV stations to operate on-channel boosters to fill in coverage gaps. Very few boosters are in operation since there are not many situations where it is possible to transmit on-channel without risking interference to the primary station signal.

Because of FCC power limitations, LPTV stations and TV translators have in the past generally used 1 W or 10 W transmitters at VHF and 1 kW transmitters at UHF. However, in 1997 the FCC replaced limitations on transmitter power output (TPO) with limitations on effective radiated power (ERP), so higher power transmitters are expected to be used by many stations in the future. Very high gain antennas are commonplace, resulting in UHF ERP levels in excess of 100 kW, although with correspondingly narrow vertical apertures.

With the elimination of TPO limits, licensees may move toward higher powered transmitters and lower gain antennas to avoid the size, wind load and narrow vertical apertures associated with high gain antennas. Coverage generally extends from five to 30 miles from the transmitter, but is highly dependent on terrain and other obstacles, and whether other interfering signals are present. ERP limits under the new rules will permit wider area coverage. The FCC protects from interference only the 62 dBu contour for low band VHF, the 68 dBu contour for high band VHF and the 74 dBu contour.

In the past, the FCC accepted applications for new LPTV stations and TV translators and for major changes in existing stations only during designated filing windows, usually one week in length and announced in advance. Minor change applications, defined as those which involve no new frequency or new service area, may be filed at any time.

Mutually exclusive applications had been subject to lotteries rather than comparative hearings in the past, but the Balanced Budget Act of 1997 repealed the FCC's authority to use lotteries. It is expected that mutually exclusive applications will be subject to competitive bidding in the future, but until the FCC completes a rule making it is not known what kind of window or other timing system will be used for filing new applications.

New full power television stations, either analog (NTSC) or digital (DTV), may displace LPTV stations and translators at any time. Displaced stations may file applications to change channel without waiting for an application filing window and therefore have priority over new applicants in applying for an available spectrum. If no new channel can be found that does not cause interference, the displaced stations must go dark. The FCC has indicated that it is unlikely to accept any applications for new LPTV or translator stations or for major changes in existing stations until it has finalized the regulations for new DTV service.

In contrast to the full power system, where applications are accepted only for channels listed in the national allotment table adopted by rule making, a LPTV station or translator may be built wherever it will not cause interference to other stations. Thus the allocation system is similar to AM radio, except that interference analyses are done solely on the basis of a special FCC computer program for NTSC-to-NTSC interference. (The FCC will also entertain special terrain shielding showings.) LPTV stations and translators are required to protect digital TV stations as well as NTSC stations, and the FCC's new digital TV rules include interference ratios for determining NTSC-to-DTV and DTV-to-NTSC interference.

LPTV stations and translators are governed by Part 74 of the FCC's Rules, although Section 74.780 lists some Part 73 rules that apply to LPTVs and translators. Operating requirements for LPTV are generally more lenient than for full-power television stations, but LPTV stations eligible for must-carry status on cable television systems are subject to all Part 73 programming requirements. LPTV stations are also subject to certain Emergency Alert System (EAS) requirements; translators may rely on EAS messages transmitted by the station they rebroadcast. Like all broadcast stations, unattended operation is permitted, subject to certain monitoring requirements for alerting an operator when equipment malfunctions.

Design considerations for LPTV stations and translators are similar to those for full power stations, except that the costs are much more modest. Equipment is smaller and less expensive and requires much less power. A transmitter site at the highest elevation is desirable, unless a lower elevation is needed to avoid causing prohibited interference to another station. Antenna design is critical, because extremely high gains and directional patterns are so common. Studio equipment may be high-grade consumer type, such as Super-VHS, when dictated by economic considerations. There are no rules currently in place to allow LPTV stations or translators to transmit digital signals, but the FCC has promised to conduct a future rule making on the subject and has issued some experimental digital transmission licenses to LPTV stations.

MMDS and ITFS—Wireless Cable

The multichannel multipoint distribution service (MMDS) and the instruction television fixed service (ITFS) are multichannel, microwave based technologies that transmit in the 2500–2690 MHz band (S-band). They therefore require an adapter (downconverter) at each receiving location in order to be received on a consumer type television receiver. Operation is omnidirectional rather than point-to-point, except where directionalization is required to avoid interference to another station. Transmitter power levels have traditionally been 10 W, but the FCC now authorizes operation at 100 W, subject to an interference showing, and most MMDS operators seek 100 W authorizations. Because of the area of spectrum in which these services operate, propagation is generally line-of-sight, and an unobstructed view of each receiving location is more important than power levels.

There are 31 MMDS and ITFS channels available, each 6 MHz wide, with one extra channel in major markets. Originally known as MDS, MMDS was a single-channel common carrier service at 2150 MHz (Channel 1), which is still available. In major markets, a second 6 MHz channel is available at 2162 MHz (Channel 2); in smaller markets, the second channel is restricted to only 4 MHz and is intended for nonvideo use (Channel 2A). The 31 channels at 2500–2690 MHz are divided into groups A through H, each with four 6 MHz channels, except for the H group, which has only three. The A and B groups are interleaved, as are the C and D, E and F and G and H groups. MMDS licenses are awarded by the FCC through an auction process. ITFS applications are filed during designated filing windows.

Originally, the A through G groups were reserved for ITFS and used by educational institutions for closed circuit instruction. The H group was allocated for private use in the operational fixed service (OFS). As entrepreneurs began developing the idea of using this spectrum for pay movies and other video services, the FCC came to recognize that the technology had the potential to compete with multichannel cable television. Thus the term *wireless cable* was born. The FCC reallocated the E and F groups, and most recently the H group, to MMDS. It also gave MMDS licensees the option of operating privately rather than as common carriers, so that the licensee of the station and the video programmer may be the same entity instead of separate ones as they were in the common carrier mode.

MMDS operators are also permitted to lease unused time on ITFS channels, which often includes up to 148 hours a week. Restrictions against holding licenses for more than one MMDS channel group have been eliminated so MMDS operators can potentially put together a relatively large number of channels to compete with wired cable service. Using a technique known as *channel mapping*, MMDS operators can automatically switch consumer receivers from one channel to another. The consumer perceives this as a full-time commercial service and is unaware that each channel is being diverted to instructional use during the 20 hours a week required by the FCC.

MMDS stations are given 45 dB protection from interference within a 15 mile radius of the transmitter, as are ITFS stations during hours when they are being leased to MMDS operators. During hours of instructional operation, ITFS stations have no fixed interference protection radius, but individual receive sites that are reported to the FCC and specified on the station license are protected regardless of their distance from the transmitter. Precise channel offset may be used to reduce co-channel protection to as little as 28 dB, but only if the licensees of both stations involved consent.

MMDS and ITFS services are based on the use of all channels, including first adjacent channels, even though the consumer type television receivers on which programming is displayed are not designed to reject strong first adjacent channel interference. Some adjacent channel protection can be built into downconverters. However, the principal technique to avoid adjacent channel interference is collocation of transmitters. The FCC's interference rules providing for a 0 dB first-adjacent channel protection ratio can best be achieved if both stations are at the same site. The FCC permits one licensee to force another to collocate at the newcomer's expense, subject to certain conditions.

In 1996, the FCC issued a declaratory ruling concerning digital transmissions in these services, and in 1997 issued the first permanent license for digital MMDS operation. Operators who wish to do so will now be able to convert to digital transmission using modified or new hardware with a choice of several modulation formats.

SYSTEM DESIGN CONSIDERATIONS

System design begins with the need for coverage of a geographical area by a television signal source. This source may consist of a single channel of NTSC video, a group of channels, a multichannel array of encoded channels or digital compressed TV signals.

Although differences exist in the propagation characteristics in VHF, UHF and S-band signals, coverage is primarily *radio line-of-sight*. Coverage is often more limited by terrain than the transmission power selected. It is generally most effective to choose the highest available site with suitable facilities and select the lowest transmitter power that will provide the required coverage. The usefulness of a particular site is often influenced by the difficulty in delivering input signals to the site, and the availability of a building, ac power and tower space. Table 6.12-1 indicates typical coverage areas achieved with the nominal output powers most commonly selected in these low power services.

Table 6.12-1 assumes minimal terrain roughness, analog transmission and a reasonable transmitter site. Conditions affecting a particular system can substantially alter the coverage areas achieved. Figures in the table were derived from FCC coverage curves and extrapolations to S-band.

Propagation Characteristics

In a TV radio frequency (RF) system, coverage beyond visible line-of-sight can be achieved by *refraction* or *diffraction*.

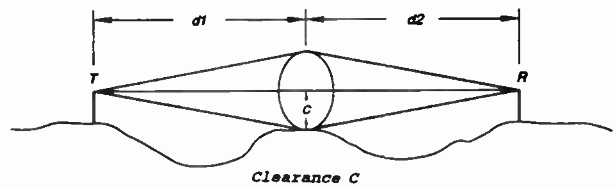
Refraction is the bending of electromagnetic waves that occurs as a signal propagates through different densities of the atmosphere. This effect causes signals to be extended beyond visible line-of-sight. While the effect is somewhat frequency dependent (in the frequency range in which these services operate), it can be approximated by considering the earth to be of a larger radius. Special graphical techniques, such as 4/3 earth radius graph paper, can be used to estimate the propagation between particular points (see Figure 6.12-1). This approach is useful in determining if sufficient clearance is available at a translator site as well as the potential for successful coverage of a particular area to be served. While variations in atmospheric conditions can modify the nominal value of refraction, the 4/3 earth radius approach is sufficient to estimate propagation in most U.S. domestic applications.

Table 6.12-1
Typical signal coverage in miles.

	Coverage in Miles			
Transmitter Power (W)	1	10	100	1,000
VHF	6	22	N/A	N/A
UHF	4	16	22	28
2.5–2.7 GHz	3	12	20	N/A

A cross section of a path between sites shows a series of concentric circles known as *Fresnel zones*. Six-tenths of the first Fresnel zone clearance is desirable to achieve propagation with near free-space path loss. The first Fresnel zone size is calculated in Equation 6.12-1.

Diffraction is an additional bending of the wave and occurs as signals approach an obstruction. This effect varies according to the smoothness of the obstructing surface relative to the wavelength of the propagating signal. The most dramatic effect is known as *knife edge* diffraction. This common phenomenon prevents visible line-of-sight into an area shadowed by a hilltop, but allows coverage of signal into the shadowed area. Signal levels in the *diffraction region* fall off rapidly. Nevertheless, it is a common method of receiving signals at a site behind a structure. Figure 6.12-2 shows the loss of signal in the diffraction as well as the results of path clearance in excess of 0.6 first Fresnel zone.



where:

$$C = 0.6F_1$$

and:

F_1 = first Fresnel Zone radius

$$F_1 = 31.6\sqrt{\frac{\lambda d_1 d_2}{d}} = 548\sqrt{\frac{d_1 d_2}{fd}}$$

d_1 = distance to near end of path (km)

d_2 = distance to far end of path (km)

d = $d_1 + d_2$ = total path length (km)

λ = wavelength (m)

f = frequency (MHz)

Equation 6.12-1. Patch clearance.

Receiver Sensitivity

A factor that may be frequency dependent is receiver sensitivity. S-band and UHF receivers are often less sensitive than VHF receivers. Receiver sensitivity can be determined by measuring system noise figure. Noise figure is defined in Equation 6.12-2.

$$\text{Noise Figure} = 10 \log_{10} (NF)$$

where:

$$NF = \frac{S/N^{input}}{S/N^{output}}$$

and:

S/N is the ratio of Signal Power to Thermal Noise Power

Equation 6.12-2. Noise figure.

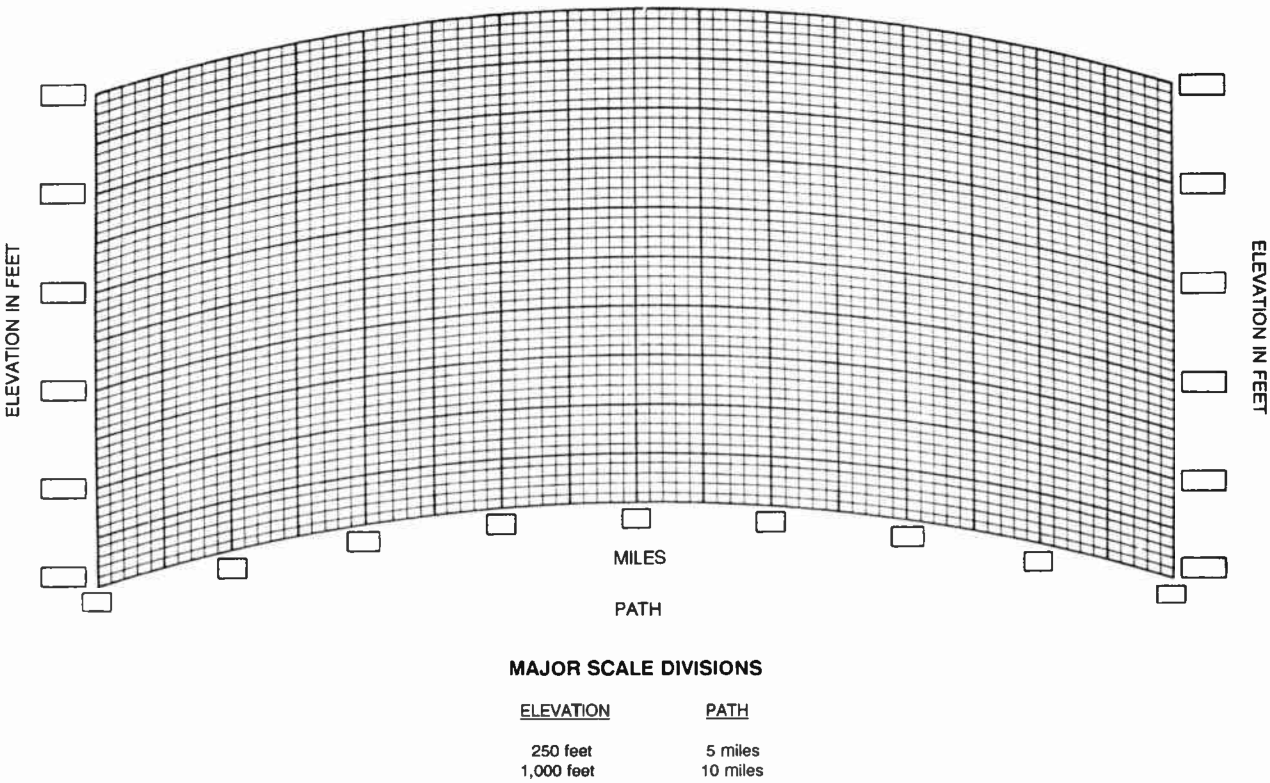


Figure 6.12-1. 4/3 earth radius path profile.

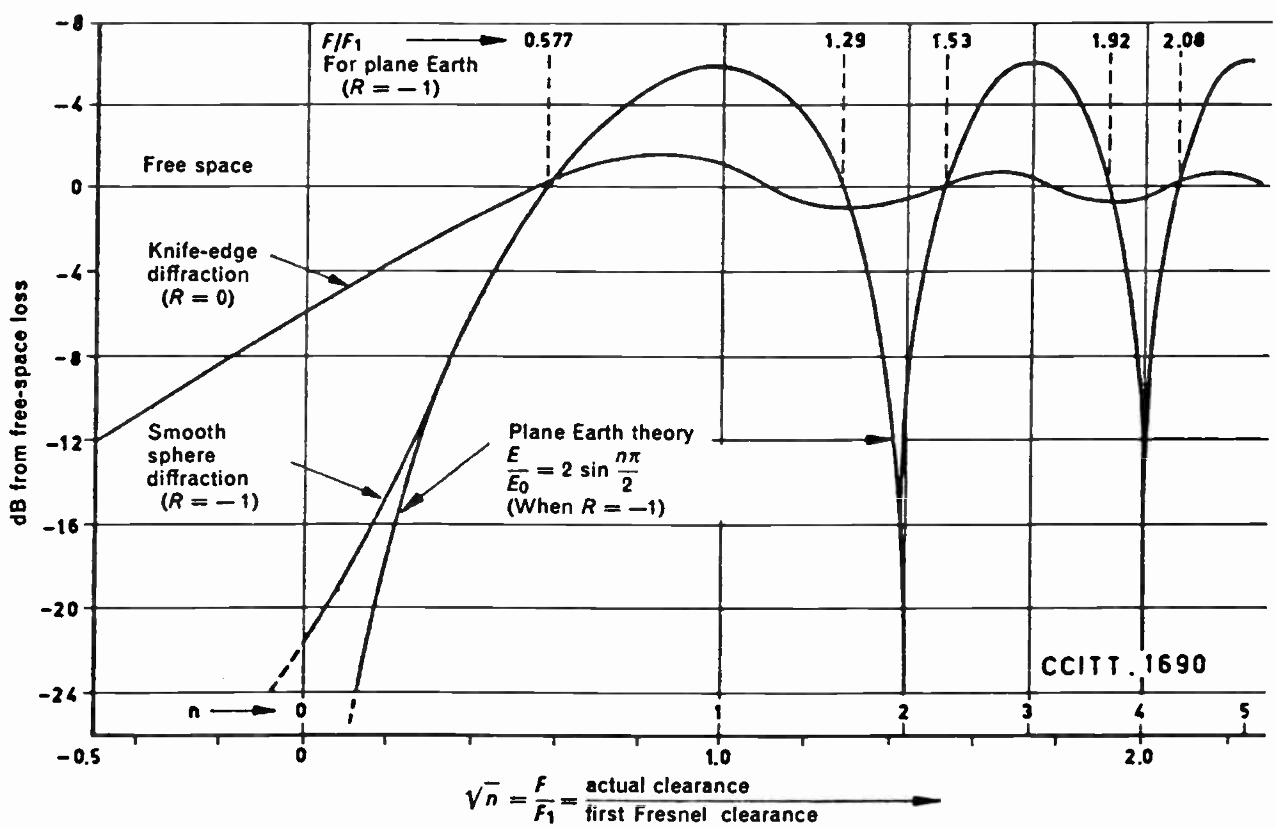


Figure 6.12-2. Path clearance.

As a signal is processed by a system, the degradation of the signal-to-noise ratio is affected by the noise contributed by the system. A low noise figure indicates that a system adds little thermal noise to the signal. Lower noise figures are more easily achieved at lower frequencies because of available electronic device technology. Recently, however, gallium arsenide field effect transistors (FETs) have become available that provide S-band noise figures comparable to the best noise figures available at VHF and with relatively little cost difference. Noise figures on the order of 1 dB to 7 dB are typical in these services.

An additional noise source often exists at VHF. In low band VHF channels, noise other than normal thermal noise may be present due to atmospheric or electrical discharges. This fact limits the usefulness of the low band VHF channels in some translator and LPTV applications. In these cases, the dominant noise source may limit the ability to receive low band VHF channels at a translator site or at ultimate viewer locations. For this reason it is customary, when possible, for high band VHF channels to be selected over low band VHF for translators and LPTV services. However, UHF channels are often preferred because of the higher power limit authorized by the FCC and the greater likelihood of finding available spectrum.

Output Power

FCC Rules establish standard power limits for the low power services. LPTV stations operating in the VHF band have an ERP limit of 3 kW. UHF stations are limited to 150 kW, operating in the NTSC mode. When digital operation begins, the limits will be 300 W at VHF and 15 kW at UHF. Circularly polarized (CP) operation is authorized and results in the doubling of transmitter power, although the availability of effective CP receive antennas is limited. In the ITFS and MMDS bands, a TPO of 10 was formerly a common nominal power rating, but 50 W and 100 W are increasingly common power levels in this service.

Boosters are often limited by system considerations. The classic on-channel repeater or booster receives an input channel or channels and simply amplifies and rebroadcasts the signal into a shadowed area with a directional antenna pattern. To avoid feedback, this requires a high gain amplifier and a high degree of isolation from the transmit-to-receive antenna. Systems of this type are generally limited to 1 W or less per channel, but operate with increased power if special shielding techniques are employed. Feeding of boosters via FM microwave eliminates this feedback problem.

Another limitation on ERP is the requirement to protect other stations from interference. The contours to which other LPTV stations and translators and ITFS/MMDS/and operational fixed service (OFS) stations must be protected are shown in Table 6.12-2. LPTV and translator stations must protect the Grade B contour of full power television stations, in accordance with ratios in Section 74.705 of the FCC's Rules.

Table 6.12-2
Protected coverage.

Service	Protected Contour or Required Separation	FCC Rules
LPTV/TRANSLATORS CH. 2–6	62 dBu Field Strength Boundary	Part 74.707 (a)(1)(i)
LPTV/TRANSLATORS CH. 7–13	68 dBu Field Strength Boundary	Part 74.707 (a)(1)(ii)
LPTV/TRANSLATORS CH. 14–69	74 dBu Field Strength Boundary	Part 74.707 (a)(1)(iii)
ITFS	50 Mile Separation	Part 74.903(b)
MMDS (OMNI PATTERN)	15 Mile Radius	Part 21.902 (d)
OFS	50 Mile Separation	Part 94.63

Channel Selection

The channel selection process involves compliance with the interference protection requirements of FCC Rules. LPTV stations and translators operate on a *secondary* basis. They are permitted to use a channel if they accept the risk that they may have to select and move to a new channel if the channel is later assigned for full power use, or if the LPTV or translator would cause interference to a full power station operating on a different channel (such as adjacent channel or UHF taboo interference).

The FCC's reallotment of TV channels 60–69 for public service use, undertaken at the direction of Congress, will also affect LPTV stations and translators since no more applications are currently being granted for new LPTV stations on these channels. Existing stations will eventually be required to relocate to lower channels unless they are able to purchase their existing channels in future spectrum auctions. The possibility of using Channels 60–69 as a transitional home for LPTV stations and translators during the DTV transition remains to be resolved by the FCC.

Because translators depend on output-to-input frequency separation to achieve isolation, it is advisable to provide several channels of separation between a translator input and translator output channel, although a minimum guardband of one empty channel may be used.

Multichannel System

Multichannel combining techniques may be employed to allow the transmission of more than one channel from a single transmit antenna. The standard technique involves using selective channel filters coupled through a directional device into a common transmission line. Important considerations involve insertion loss per channel, power handling capability, channel combiner size/cost and antenna system bandwidth. Channel combiners for the ITFS/MMDS band are available that efficiently couple up to 16 alternate channels onto a single transmission line. The transmission of adjacent channels can be accomplished through the use of separate transmit antenna systems, broad band hybrid couplers or special filters.

Antennas

Transmit antennas often concentrate power in a desired direction. Calculated as antenna gain, this concentration of power is a measure of the effectiveness of radiation of an antenna compared to a simple dipole antenna or a point source radiator. Two important parameters determine antenna gain: *vertical beam width* and *horizontal beam width*.

Power is concentrated vertically in a narrow beam by the use of stacked radiators. The signal emanating from this array adds in-phase in the direction of the main beam and cancels in other directions, effectively increasing the power transmitted in the main radiation pattern of the antenna. The azimuth pattern of the antenna is determined by the placement of the radiating elements, which are normally arrays of dipoles or of slots in a waveguide. Both types of antennas are widely used from VHF through S-band.

While a variety of antenna patterns are available, the most common patterns are the *omnidirectional* and *cardioid*. The omnidirectional attempts to radiate power equally in all horizontal directions. A cardioid pattern is shown in Figure 6.12-3.

This basic pattern and its variations are available with antenna gains from 10 dB to 18 dB, depending on the number of bays of radiating elements utilized. Vertical beam width narrows with increased horizontal gain as seen in Figure 6.12-4.

Additional considerations in antenna selection and specifications are *beam tilt* and *null fill*. Beam tilt involves radiating the horizontal pattern of the antenna slightly away from the horizon. Beam tilt provides coverage into terrain below the antenna, an objective commonly accomplished by specifying − 1/2° or − 1° of beam tilt. Null fill provides a small amount of radiation in the direction of an antenna null to provide some degree of coverage in the area of signal cancellation.

Directional antennas constructed with parabolic or

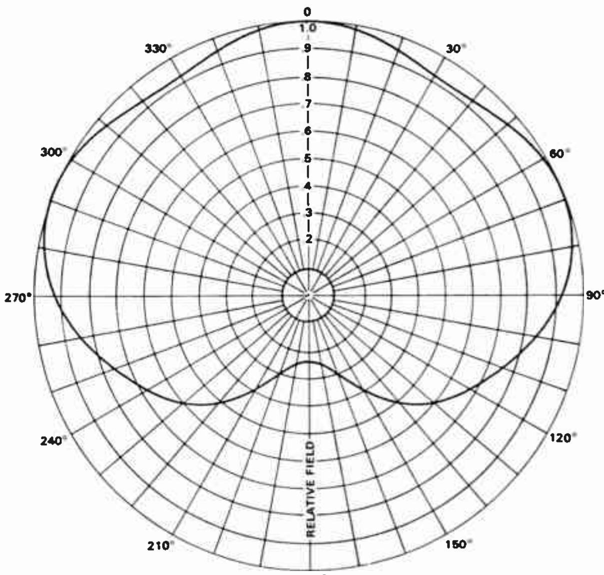


Figure 6.12-3. Cardioid antenna pattern. (Courtesy Andrew Corporation.)

cylindrical reflectors are occasionally used for transmission but are commonly used for receive purposes at repeater sites and at ITFS/MMDS receive locations. This type of antenna is very directional compared to standard transmitting antennas, and gains above 20 dB are common. Horizontal beam widths less than 6° can easily be achieved. A common receive antenna with specifications is shown in Figure 6.12-5.

Antennas have limitations in power handling and bandwidth, although antennas with special capabilities are manufactured and sold at higher costs.

Site Requirements

Transmitter sites are normally located on hilltops, in tall buildings or at convenient tower locations. In addition to the obvious requirements for receiving and transmission hardware, a site should be selected to provide the desired coverage with the available antenna mast and selected antenna. Site requirements encompass three issues: environmental considerations, ac power requirements and grounding and transient protection.

Environmental Considerations

Electronic equipment is designed to operate over a normal range of temperatures and within a reasonable range of moisture and dust conditions. Maintaining appropriate room temperatures and equipment cleanliness improves the reliability of a system. A good rule of thumb concerning equipment reliability is the common rule applied to chemical reactions: most reactions double in speed for each 10°C (50°F) rise in temperature. Since many detrimental environmental effects, such as contact corrosion, are chemical reactions by nature, the speed of chemical reactions will affect equipment longevity. Corrosion, metal migration in transistors and the degradation of electrolytic capacitors closely follow this general rule.

Maintaining low temperatures at a site will usually enhance reliability. Most equipment operates comfortably in ambient temperatures as low as 0°C (32°F). Temperatures above 35°C (95°F) will normally be outside the range the equipment was originally tested in and should be avoided.

It is advisable to provide filtered air to a building and a method of easily removing and cleaning the building air intake filters. The cleanliness of air filters and metal heatsink surfaces will affect cooling system efficiency. Most air filters are washable, and a supply of water with detergent should be available at a site. If equipment is exhausted directly to the outside, provide a substantial supply of unrestricted building intake air. Most transmitter cooling systems are not designed to overcome substantial air pressure drops.

Moisture is another site consideration. Equipment normally operates above room temperature, making condensation of water on the equipment unlikely. However, in some instances it may be necessary to provide additional site heating to keep the equipment dry. It may also be desirable to prefilter or route inlet air through baffles or ductwork to eliminate water drop-

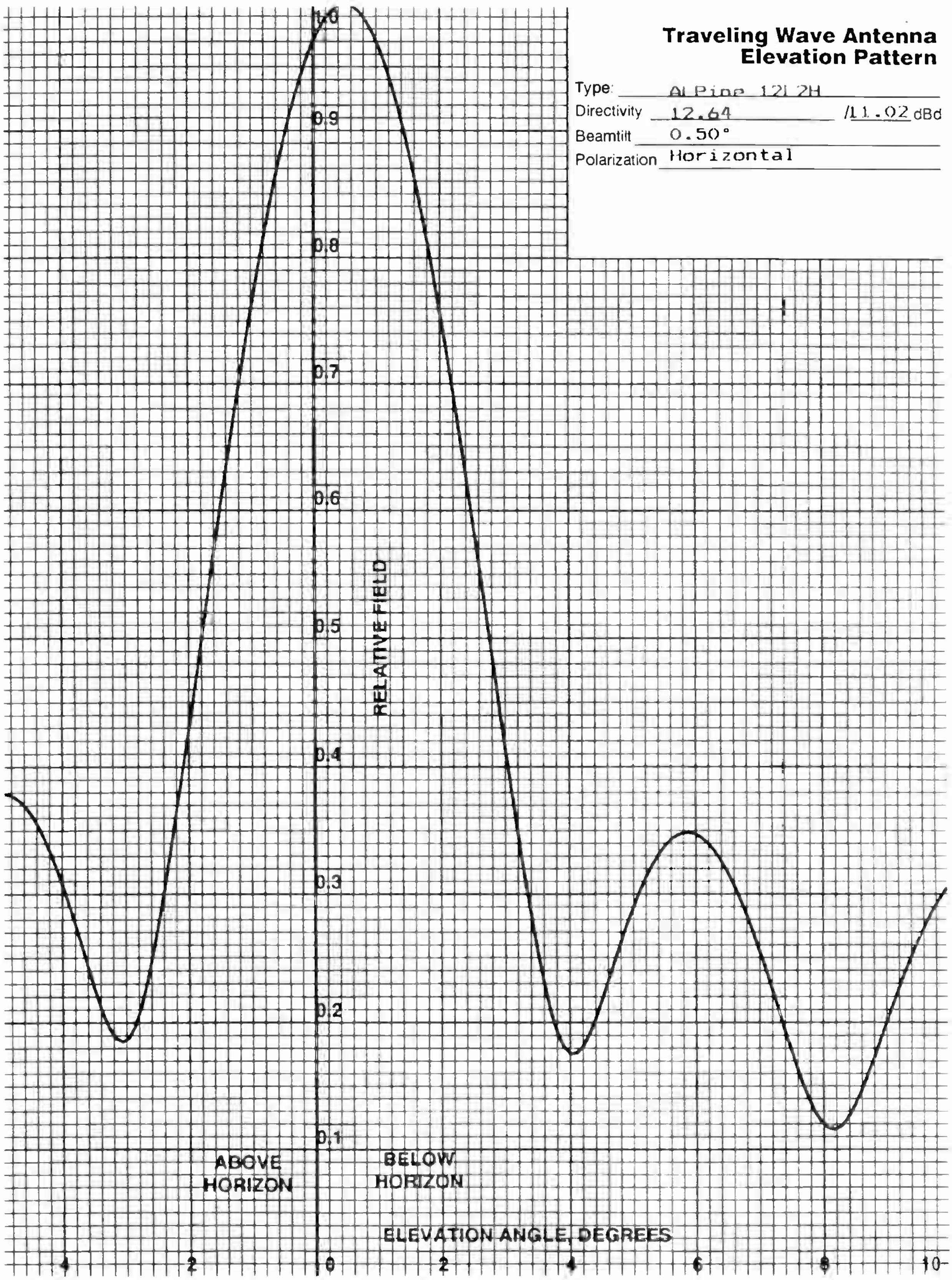
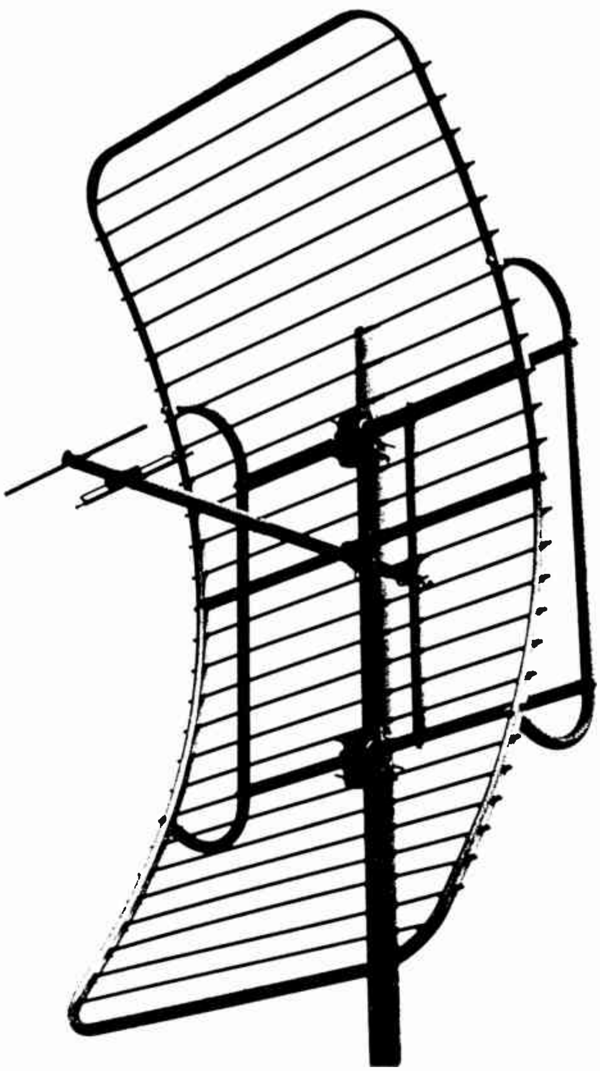


Figure 6.12-4. Twelve-bay antenna vertical beam width (Courtesy Andrew Corporation.)



Cylindrical Parabolic Grid Reflector Antenna

Band: UHF
Gain: 18 dB nominal
Front-to-Back Ratio: 20 dB minimum
VSWR: 1.2:1
Power Rating: 100 Watts

Figure 6.12-5. Directional UHF antenna. (Courtesy Scala Electronics Corporation.)

lets. Exhaust air should be directed away from the direction of prevailing wind. Although exhaust air screens necessary to prevent birds and animals from entering the building, if not regularly cleaned the screens may become clogged and adversely affect equipment cooling. Automatic motor-controlled louver systems can be used to allow the equipment to maintain a reasonable room temperature in cold weather. However, if the motorized louver system malfunctions, it can result in equipment overheating. Louver systems that rely on air pressure to open must be used with caution, since additional pressure drop is incurred.

AC Power Requirements

Site requirements for power must take into account transmitter demands, cooling/heating requirements, test equipment and lighting needs. A 1 kW UHF transmitter typically draws more than 25 A of ac current at 230 ac volts. A 1 kW UHF transmitter will produce about 4,000 W of exhaust heat that must be vented from the room or cooled with an air conditioning system. This converts to 13,600 Btu per hour of air conditioning requirement.

In a multichannel system it is advisable to provide an ac shut-off per cabinet. This permits power to be shut down in a particular cabinet so maintenance can take place while the other cabinets continue to operate. Wall circuit breakers are thermally operated devices and should be sized substantially above the equipment requirements to allow for in-rush transients and thermal heating of the building. It is common to run conduit directly to cabinets for higher power systems. Lower power transmitters can often be plugged into standard ac outlet strips.

Grounding and Transient Protection

Transmitter sites require an effective earth ground to minimize ac hum problems, provide ac line transient protection and allow a safe path for lightning induced current. A site ground should bond the ground provided from the power company with a grounding rod associated with the tower. A copper-clad steel grounding rod buried in the earth and connected to the tower, building structure and ac power line ground is standard practice. It is also advisable to add a ground strap, often constructed with flat copper strapping or heavy copper braid, from equipment cabinets to the site ground. Transmission grounding kits, which are available for all standard transmission line cables, provide a low resistance method of connecting the ground system to the main transmission line.

Transient protection is necessary on both the ac line and RF transmission lines. Most modern transmitters incorporate some form of internal transient protection. The most common method is the installation of transient devices across the ac line and from each ac line connection to chassis ground. These transient protectors are often of the metal oxide varistor (MOV) type and are available in various voltage ratings. Because of the wide variation of power line voltages often present at transmitter sites, choose the highest rating varistor that will provide the required transient protection. For example, it is often desirable to use a 150 V MOV instead of a 135 V MOV to minimize transient protector damage from a momentary high line condition. The higher voltage MOV will typically provide the necessary momentary transient protection. By connecting MOVs from line-to-line and line-to-chassis ground, common mode and unbalanced transients are clipped. Large MOVs can be installed at the ac power box to provide additional protection. More sophisticated transient protectors that insert inductive elements in series with the ac line prior to transient protection are also available.

RF input and RF output transmission lines should incorporate proper grounding kits on the outside shield of the cable. Direct current return paths should be provided in receiver and transmitter circuitry to reduce the danger of damage from static buildup of the RF transmission line center conductor. Harmonic filters that employ shorted quarter-wave trap sections provide a low resistance dc return path to ground and further enhance the RF transmission line transient protection.

Long transmission lines that run between a transmitter site and a tower can result in lightning induced transient damage. Improved bonding between the site and the tower may be necessary. Enhanced transient capability of the receiver or transmitter circuit through the use of spark gap capacitors and low voltage transient protectors can also enhance reliability. Contact the equipment manufacturer to solve specific transient damage problems.

HARDWARE ANALYSIS

Describing the range of hardware required to cover VHF through S-band requires a general approach. Most television transmitters can be divided into three sections: modulator, upconverter and final amplifier. The modulator processes incoming video and audio and performs the basic modulation process at intermediate frequency (IF). The upconverter heterodyne converts the signals to the desired output frequency and provides additional processing such as level control and precorrection. The final amplifier raises the signal level to the desired output power. Peripheral equipment often includes encoding and multichannel combining equipment.

Heterodyne translators are similar except that the modulator is replaced by a receiver assembly that converts the input signal to IF. Boosters have two standard configurations: a simple on-channel amplifier and a video-fed transmitter.

A full RF transmission system requires suitable transmission lines, filters, antennas and other monitoring components. Receive hardware consists of down-conversion equipment and decoding hardware.

Transmitters—NTSC

A modulator performs the function of producing a vestigial sideband amplitude-modulated picture carrier and a frequency-modulated aural carrier at convenient intermediate frequencies. Since signal level during demodulation determines picture brightness, accurate modulation levels that are independent of average picture levels must be maintained. This is accomplished with a video clamp circuit that holds either sync tip or black level to a specific voltage level. Other video processing, such as differential phase correction and video sync stretch, can also be accomplished prior to modulation. It is also desirable to add white clipping and sync spike clipping prior to modulation to prevent overmodulation or incorrect output levels at the output of the modulator. Vestigial sideband filtering is achieved with a suitable surface acoustic wave (SAW) filter after modulation. These filters provide stable, accurate characteristics with sharp roll-off and reasonably flat delay variation across the band.

Aural modulation is usually frequency modulation of an IF carrier. Standard IF frequencies chosen in most modulators are 45.75 MHz for visual and 41.25 MHz for aural. These frequencies correspond to the standard IF of television sets and other common test equipment such as demodulators. With the advent of Broadcast Television Systems Committee (BTSC) stereo television, it became desirable to have an FM modulator capable of wide deviation with low distortion. It also became desirable to phase lock the center frequency of the FM carrier to the visual IF oscillator to maintain accurate visual/aural frequency separation. This can be accomplished with a very narrow band phase-locked loop that maintains the average center frequency of the FM carrier. FM modulators normally have a preemphasized audio input for standard baseband audio, and a wideband unbalanced input without preemphasis for composite audio input signals. Some modulators provide a separate filtered input for incorporating high-frequency subcarriers.

Figure 6.12-6 illustrates the functional block diagram of a typical modulator/upconverter amplifier assembly. Intermediate frequency processing is performed to linearize the RF characteristics of the output amplifier array. This linearization is a combination of amplitude and phase precorrection. Amplitude precorrection is accomplished with attenuators that are bypassed at specific signal levels by the cut-in of biased diodes. This can be achieved with little phase variation, and the sharpness of the amplitude response characteristic can be easily adjusted to match the opposite characteristics of the amplifier.

Phase correction or, more accurately, incidental carrier phase modulation (ICPM) correction can be accomplished by demodulating a sample of the IF visual carrier. The demodulated video is used to drive a phase modulator in line with the IF signal. By adjusting the amplitude and shape of the video sign, accurate compensation for phase errors that occur in the transmitter's output amplifier can be provided.

Automatic level control is an additional function that usually takes place at IF. A control loop maintains a constant IF or RF output level, and drift is compensated by adjusting a voltage-controlled attenuator at IF.

Upconversion occurs in a double balanced mixer at low level. Filtering and amplification raise the level to the amplitude necessary for detection and metering. The last section of most transmitters is final amplification. Modern transistor designs can produce 1 kW of output in UHF and 100 W of output in S-band.

Transmitters—Digital

The transmission of high bit rate information in a fixed 6 MHz channel requires an efficient modulation system. Quadrature amplitude modulation (QAM) and vestigial sideband amplitude modulation (VSB) can provide the required efficiencies. Both techniques have been well developed or are currently in an advanced

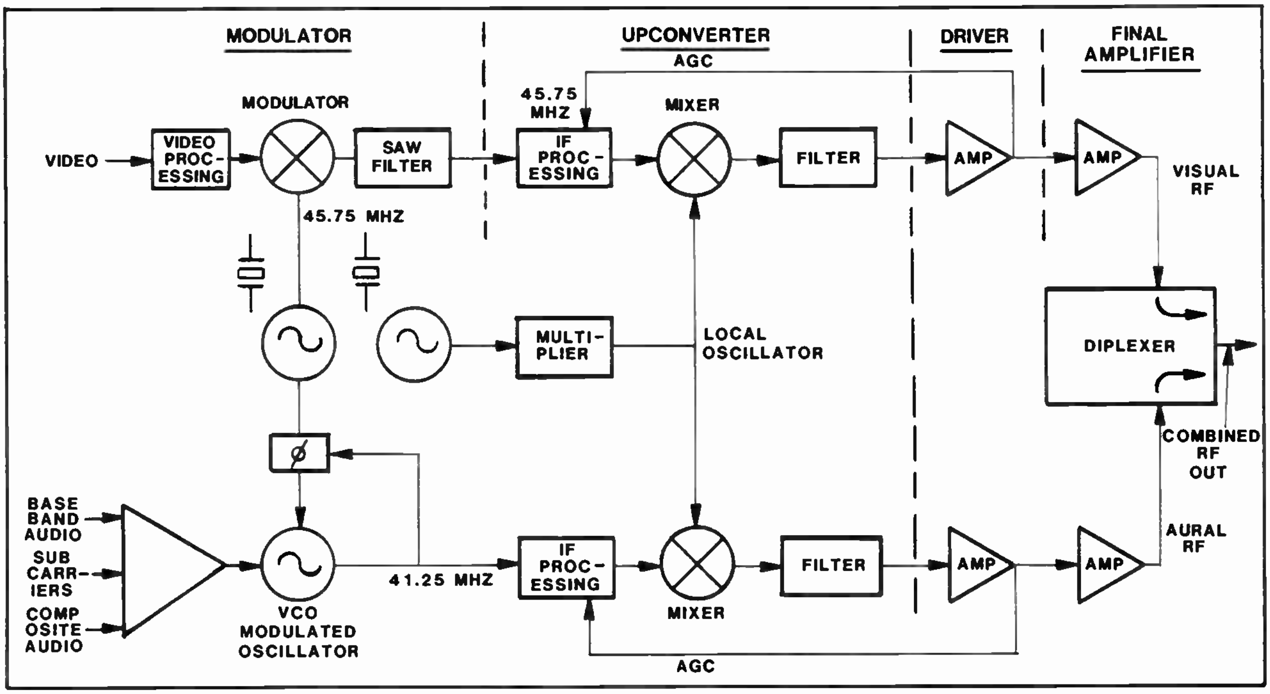


Figure 6.12-6. Television transmitter block diagram.

stage of testing. Transmitters for QAM and VSB combine IF digital modulators with upconverters and amplifiers modified or designed for digital signals.

LPTV, MMDS and ITFS services can use these digital formats in accordance with regulations in place or pending for each service. The DTV Standard adopted by the FCC for full power broadcast television is an 8VSB system. While translators and most LPTV systems will likely use the same system, it is not yet known whether any operators will seek or be authorized by the FCC to use QAM for subscription or other specialized services.

Translators

A heterodyne translator can be configured by replacing the modulator with a suitable UHF or VHF receiver. By using the same IF and operating level of the modulator, this change can be accomplished simply. The receiver needs to provide the necessary gain and filtering to avoid interference from other channels as well as avoid interference from the translator output. Adding a preamplifier at the receive antenna is a common way to reduce the system noise figure to the lowest possible level.

SAW filters are normally used in modern translators to provide filtering of adjacent channel products. As in transmitters, SAW filters provide extremely sharp selectivity with relatively constant delay characteristics. When the output amplifier is linear enough to allow operation with multiplexed IF signal, SAW filter bandwidth is chosen to pass the visual and aural carriers. For externally diplexed output amplifiers, it is necessary to provide a signal separation function in the receiver. Receivers have been developed to split the visual and aural IF carriers to allow separate processing in the following upconverter and amplifier assemblies. The splitting of the aural carrier is complicated by the fact that the incoming signal may arrive with a carrier frequency error. In addition, the incoming signal may contain stereo or other subcarrier information. These problems can be resolved with a phase-locked loop that tracks the incoming frequency and aural IF SAW filters that allow the full aural spectrum to pass.

Translator receivers require a very wide automatic gain control range. This allows receivers to be used at sites with very small input signals as well as large input signals. Some designs incorporate switchable attenuators that adjust the automatic control range of the receiver to the expected range of the input signal.

Digital translators generally will utilize the heterodyne translation technique previously described with certain modifications to filtering, power control and electrical parameters as required by the particular digital modulation format.

Boosters

Boosters can be licensed as microwave fed transmitters. In this configuration, the hardware is essentially identical to video fed transmitters.

The earliest boosters predate television translator service. The first documented boosters were installed in 1948 in Oregon. These boosters, and the others that followed were simple on-channel amplifiers connected to receive and transmit antennas. In this configuration, the system is stable and provides useful signal coverage

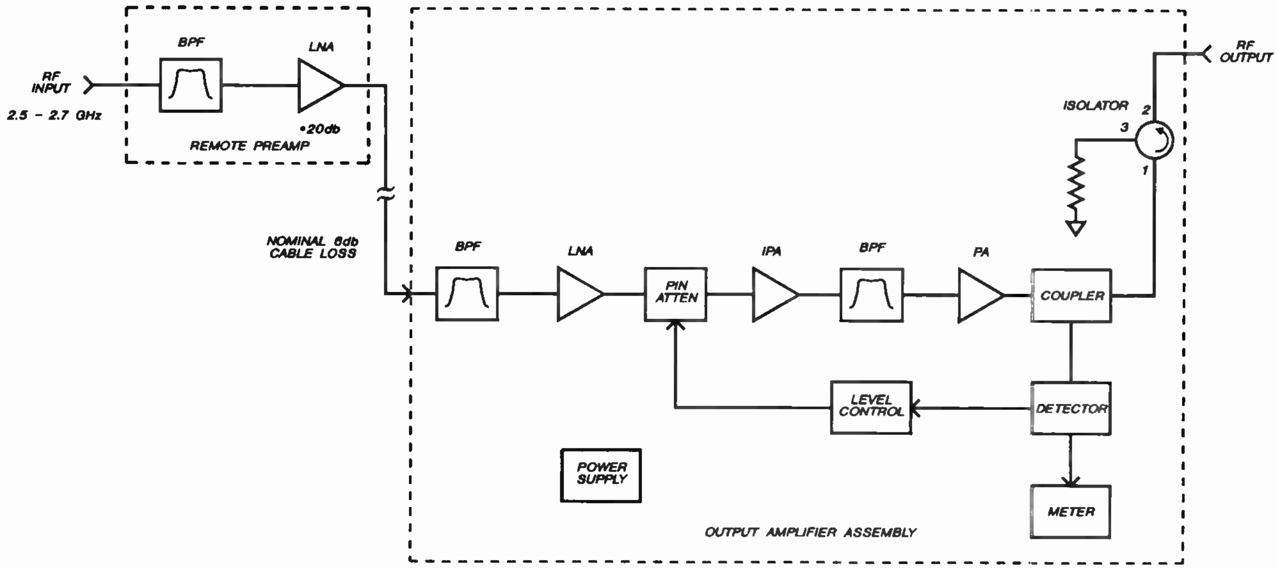


Figure 6.12-7(a). Multichannel booster block diagram.

if the antenna isolation substantially exceeds the amplifier gain. Common booster gains typically begin at 40 dB. With a directional transmitting antenna, an area within several miles of the booster site can be served. Multichannel boosters have been used in the ITFS/ MMDS band for some years. The functional block diagram of an advanced booster design is shown in Figure 6.12-7.

In this design, a remote preamplifier is mounted at the receiver antenna. A coaxial cable connects the preamplifier to the final amplifier assembly, which allows physical separation of the gain blocks, improves the stability of the system and reduces the system noise figure. The final amplifier includes a power detection circuit and automatically controlled attenuator. This feedback control loop adjusts the gain of the amplifier to maintain the output power within the capabilities of the output amplifier. Selection of the proper filters in this design enables it to work in a narrow band or multichannel environment. Open-loop gain exceeds 80 dB, and the output amplifier has up to 10 W of peak envelope power capability. This allows relatively high power levels to be transmitted. Pending booster regulation changes are expected to simplify the licensing of boosters. In other countries, boosters are performing an important function of both filling in shadowed areas and extending the coverage of low power systems.

Transmission Lines, Filters and Antennas

Both coaxial and waveguide transmission lines can be used in S-band. At VHF and UHF frequencies, coaxial transmission lines vary in diameter from 1/2 in. to 3 1/8 in. The efficiency of transmission lines varies with frequency and size. The insertion loss of various types of coaxial cable at different frequencies is shown in Table 6.12-3.

Table 6.12-3
Transmission line losses.

	Approximate Loss Per 100 Feet in Decibels				
Band	1/2" Foam	7/8" Foam	1 5/8" Air	3" Air	Elliptical Waveguide
LOW-BAND VHF	0.5	0.25	0.15	0.1	N/A
HIGH-BAND VHF	2	1	0.6	0.4	N/A
UHF	1	0.5	0.3	0.2	N/A
S-BAND	4	2.5	1.25	N/A	0.4

Both elliptical or rectangular waveguide can be used at S-band. Waveguide losses are substantially lower than coaxial transmission line losses, but the cost of waveguide is higher. Flexible elliptical waveguides are often used to feed transmit antennas in this frequency range. (EW17 and EW20, manufactured by Andrew Corporation, are typical elliptical waveguides for this frequency range.) Rectangular waveguide is widely used in channel combining networks to minimize losses. (WR340 and WR284 are common rectangular waveguide sizes for this frequency range.) Coaxial components in a channel combiner and transmission line network can be eliminated by using suitable waveguide components, such as E-plane and H-plane elbows, directional couplers and filters.

A harmonic and trap filter assembly feeding the transmission line is often used in the VHF and UHF frequency range. In the S-band, it is also common to include a multichannel combining network. Notch filters are often required in low level diplexed assemblies to trap out-of-band spurious emissions.

A multichannel environment requires several channels to be coupled onto a common transmission line. This can be achieved by using a multiplexing filter. While this filter can take several forms, a typical directional filter allows a specific channel to be directionally coupled to the output and rejects signals of other fre-

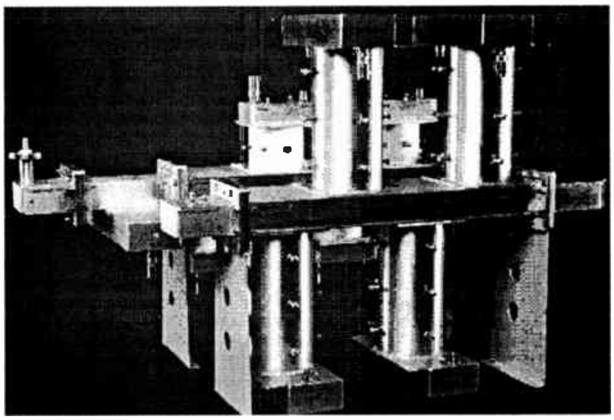


Figure 6.12-8. Typical directional filter.

quencies that pass over the filter unattenuated (see Figure 6.12-8). A diagram of a multichannel directional filter is shown in Figure 6.12-9.

Transmitting antennas are normally of the slot or stacked dipole design. A typical LPTV antenna with directive elements attached is shown in Figure 6.12-10. Side mounting can affect the pattern of a transmitting antenna. Fortunately, computer programs are available that can predict the impact on the pattern and allow this effect to be addressed in system design (see Figure 6.12-11).

S-Band Receiving Equipment

A typical receiving system for S-band is shown in Figure 6.12-12. In this system, a directional receive antenna couples signal into a broad band block downconverter, which heterodyne converts the signal to VHF superband. The signal is then decoded and remodulated onto a low band VHF channel.

Digital service is a more complex but similar technique of extracting the desired program material from the received input and converting it into a standard NTSC signal for television viewing.

The Receive Antenna

Receive antennas vary from 12 dB gain corner reflectors to 36 dB gain parabolic dish designs. A short

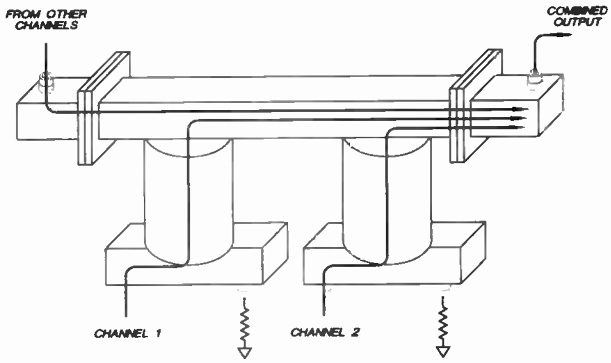


Figure 6.12-9. Two channel combiner.

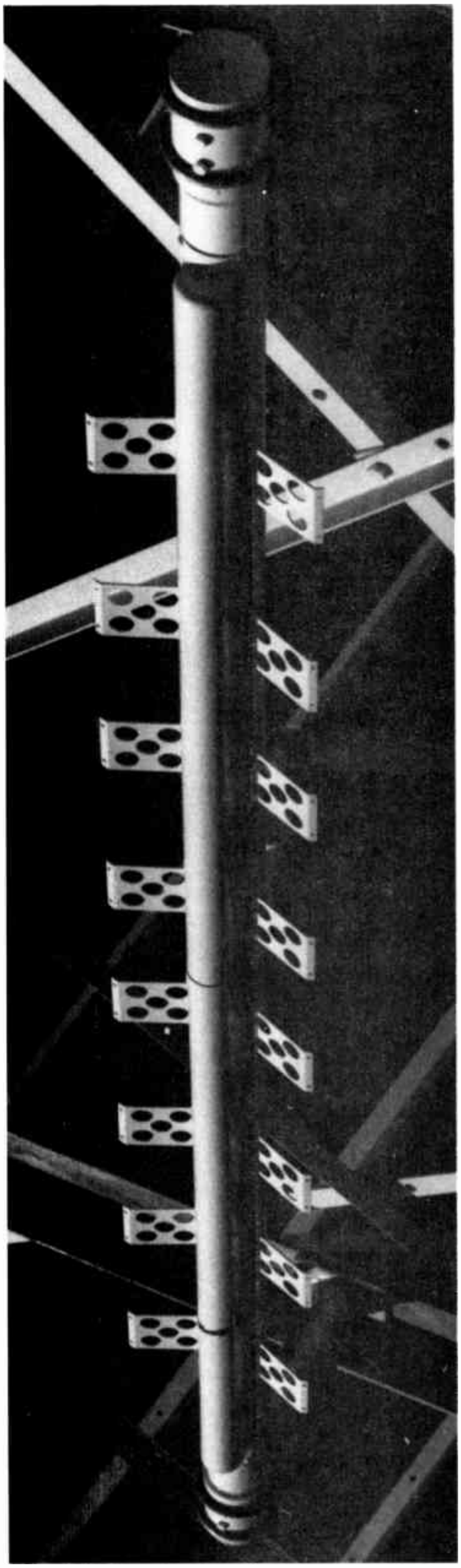


Figure 6.12-10. UHF slot array transmitting antenna. (Courtesy Andrew Corporation.)

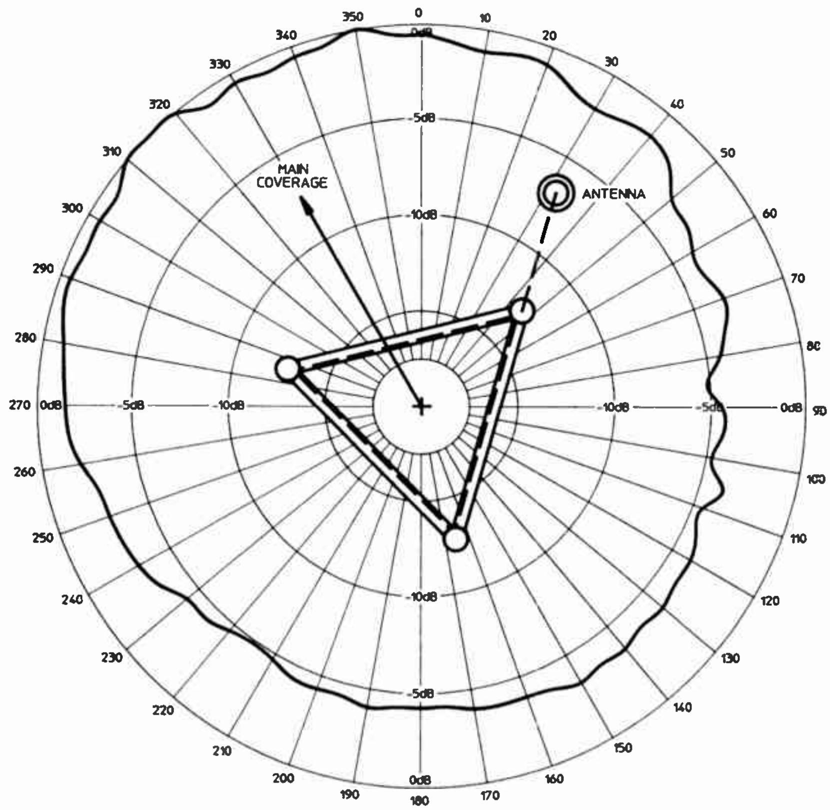


Figure 6.12-11. Calculated pattern of side mounted antenna. (Courtesy Dielectric Communications.)

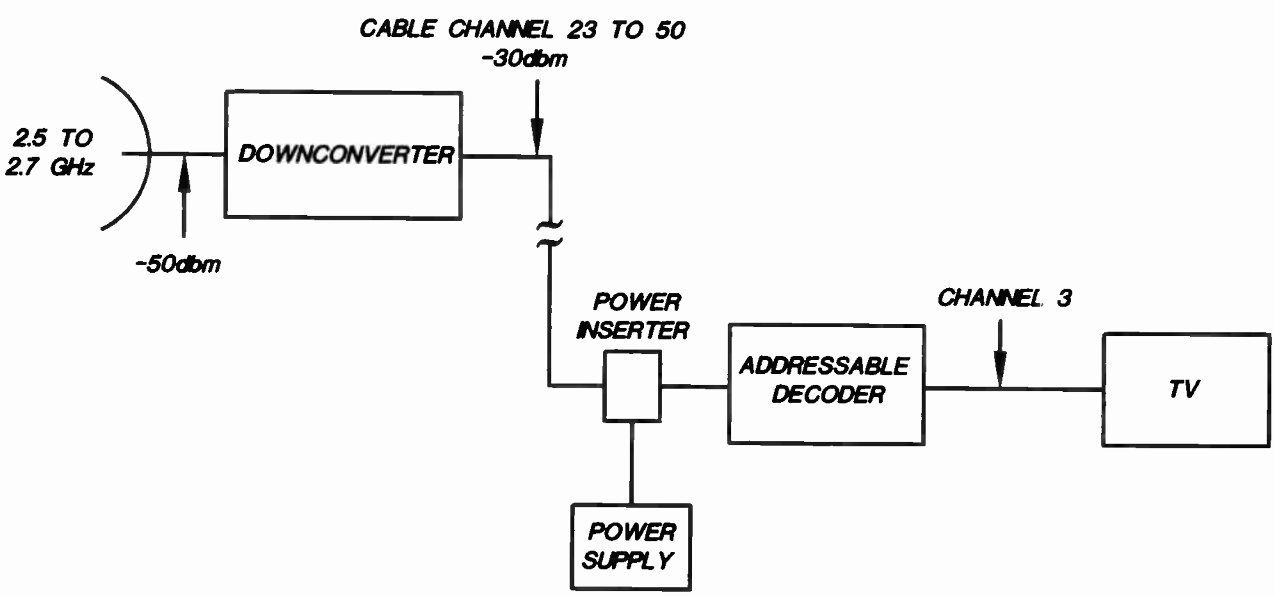


Figure 6.12-12. S-band receiver.

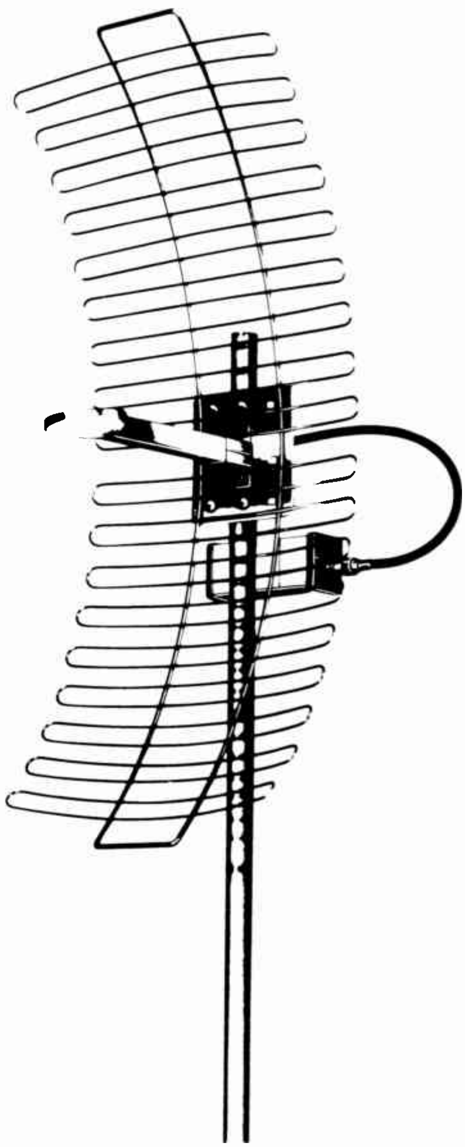


Figure 6.12-13. 21 dB gain ITFS/MMDS receiving antenna. (Courtesy Lance Industries.)

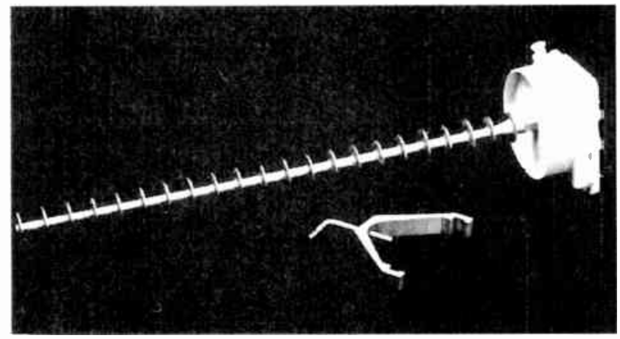


Figure 6.12-14. ITFS/MMDS downconverter. (Courtesy California Amplifier.)

jumper usually connects the antenna feed directly to the downconverter, although in some designs the receive feedhorn is directly mounted to the downconverter. A typical antenna is shown in Figure 6.12-13.

Downconverters

The key characteristics of downconverters are sensitivity, signal handling capability, gain, spurious products, phase stability and reliability (see Figure 6.12-14 for a typical example).

Sensitivity is achieved through the use of low noise transistor stages preceding the mixer. Noise figures of 4 dB are typical.

Signal handling capability is a limitation in a multichannel environment. The front-end mixer and the VHF output amplifier are both sources of distortion, which takes the form of composite triple beat, a common cable television distortion characteristic. Distortion can be minimized through improved mixer design and increased VHF amplifier output capability. It can be tested with a two-tone input test signal with the objective of achieving -27 dBm per tone at the input of the converter without significant output distortion.

The gain of a downconverter is generally not critical unless there are long transmission line feeders after the converter. A normal range for converters is 17 dB to 30 dB. Downconverters with as much as 40 dB of gain are available, but these units may suffer from output signal distortion because of high signal levels in the output amplifier.

Spurious products are a significant problem in many cases. Downconverters use various local oscillator techniques to achieve clean local oscillator signals, including phase-locked loops, dielectric resonators, free-running voltage-controlled oscillators (VCO) and SAW oscillators. Spurious products in the output band should be suppressed at least 50 dB below normal output levels.

Phase jitter is a result of local oscillator electrical noise and mechanical vibration sensitivity. This characteristic leads to horizontal streaking in the picture and difficulty in phase lock demodulating television signals. Especially important in digital systems, phase jitter can be observed on a spectrum analyzer tuned to an output unmodulated carrier.

Decoders—Analog

Decoders typically have synthesized front-ends and specialized IF decoding circuitry. Addressability is a desirable characteristic of decoders but adds to the cost of the unit. Decoders designed for cable television service will occasionally have difficulty with the widely varying levels in an MMDS system and may be unable to handle the normal adjacent channel variations that it experiences. The time constant of an automatic gain control of the decoder may also be inappropriate for an MMDS signal environment. As MMDS systems grow, it is expected that new decoders developed specifically for this service will become available.

Since decoders heterodyne convert or remodulate the signal onto a low band output VHF carrier, additional video and audio signal degradation may be incurred.

Decoders—Digital

In addition to the RF and IF signal processing present in analog decoders, digital units must demodulate a specific 6 MHz channel, demultiplex the desired program, decompress the bit stream and ultimately reconfigure the signal into NTSC format video and audio. Powerful large-scale integrated circuit chips are now available to perform many of these functions, and commercially available digital decoders are now available for several digital modulation standards.

6.13
CABLE TELEVISION SYSTEMS

WALTER S. CICIORA, Ph.D.
SOUTHPORT, CT

INTRODUCTION

According to A. C. Nielson, cable television service is enjoyed by more than 65 million U.S. households, market penetration of over 67%. Cable service passes 96% of U.S. households.

Because cable television has been so successful and has enjoyed such vigorous growth and acceptance, it has spawned competitive technologies, including pre-recorded media (pre-recorded tape and disk sales and rentals) and direct broadcast satellite (DBS), as well as the interest of the telephone industry. Its high visibility has also attracted the attention of regulators and legislators since important public policy issues are involved. Understanding what cable television is, how it works and its economics will help decision makers in these arenas. This understanding will also aid technologists in determining which technologies are appropriate for cable television applications and which are not. For some readers, the main interest in this chapter is simply a better understanding of the technology that supplies their home with video. This chapter is intended for technical professeionals who require a briefing on cable television in the United States.[1]

Historical Perspective

Cable television is an important part of the way in which the citizens of the United States are informed and entertained. It provides large numbers of television channels to more than half of the U.S. population in a cost-effective way.

Prior to the 1990s, cable television systems were not intended to be general purpose communications mechanisms. Their primary and often sole purpose was the transportation of a variety of entertainment television signals to subscribers. Thus, they needed to be one-way transmission paths from a central location, called a headend, to each subscriber's home, delivering essentially the same signals to each subscriber. Then as now, the signals were intended for use with the consumer electronics equipment that subscribers already owned. This equipment was built to operate on the then current U.S. television technical standard called the National Television Systems Committee (*NTSC*) after the organization that created it in 1941. This black-and-white television standard was modified in 1953 to provide compatible color information to color television receivers and again in 1984 to add compatible stereo sound.

Cable television's original purpose was to deliver broadcast signals in areas where they were not received in an acceptable manner with an antenna. These systems were called community antenna television (CATV). In 1948, Ed Parson of Astoria, OR, built the first CATV system consisting of twin-lead transmission wire strung from housetop to housetop. In 1950, Bob Tarlton built a system in Lansford, PA, using coaxial cable on utility poles under a franchise from the city and Jerrold Corporation's antenna booster amplifier. Martin Malarkey opened his cable system in Pottsville, PA, within a week of Tarlton. Malarkey used Radio Corporation of America (RCA) Antennaplex equipment, which was originally developed for the in-house distribution of the National Broadcasting Company's (NBC's) signals in the RCA building in New York City. Malarkey also organized and developed the National Community Television Association (NCTA), which later was renamed the National Cable Television Association.

In most CATV systems, off-air signals were not available or were very weak because of the terrain or the distance from television transmitters. In some areas, such as New York City, multiple signal reflections and shadows cast by buildings made reception difficult. In both of these environments, a hard-wire method of delivery of signals to subscribers was welcomed. The first operators of these systems were retail TV receiver dealers who sought to expand the market for the sale of their products by also providing the signals the products required. By the late 1960s, nearly all of the areas of the United States that could benefit from a community antenna had been served. Growth in the cable industry all but stopped.

In the mid 1970s, an embryonic technology breathed new life into cable television. This technology was satellite delivery of signals to cable systems, which added more channels than were available from terrestrial broadcasters. While satellites and earth stations were very expensive investments, these programming

[1] It is based on a CableLabs' publications by this author titled, "Cable Television in the United States, An Overview," which is available on CableLabs' Website: www.cablelabs.com/publications.html See also: Modern Cable Television Technology: Video, Voice, and Data Communications by W. Ciciora, J. Farmer, D. Large, Morgan Kaufmann Publishers, 1999. Also, see www.ncta.com/directory.html for the latest cable statistics.

pioneers understood that the costs could be spread over many cable operators who, in turn, could serve many subscribers.

Three categories of signals came into existence, all of which sparked new excitement in the business:

- *Super stations*—local stations that were distributed nationally over satellite and became mini-networks. (The Turner Broadcasting System of Atlanta, GA pioneered the concept)
- Specialized channels for news, sports, weather, education, shopping
- Movie channels such as Home Box Office (HBO). (Cable television became much more than just a community antenna for areas with poor reception: it became a means of receiving programming otherwise unavailable.)

Today, cable subscribers are offered a variety of video services. The foundation service required of all subscribers is called *basic* and includes off-air channels, some distant channels and some satellite-delivered programs. (The satellite programs include the super stations and some of the specialty channels.) *Pay television* constitutes premium channels, usually with movies and some special events, which are offered as optional channels for an extra monthly fee. Some cable systems offer *pay-per-view* (PPV) programming, which is marketed on a program-by-program basis. Recent movies and special sports events are the mainstay of PPV programming. *Impulse pay per view* (IPPV) allows the subscriber to order the program spontaneously, even after it has begun. The ordering mechanism usually involves an automated telephone link or, occasionally, two-way cable.

Ways of providing conditional access to allow for a limited selection of service packages at differing price points are often included in the cable system. Simple filters remove the unsubscribed channels in some systems, while elaborate video and audio scrambling mechanisms involving encryption are used in other cable systems.

During the early-to-mid-1980s, a wide variety of other services were offered to cable subscribers: videotext, teletext, other forms of electronic publishing and information age services, home security and digital audio programming. Subscribers in the 1980s showed a remarkable lack of interest in these services. It is interesting to speculate on the reasons for this disinterest, but the negative experience was without exception. While subscribers were enthusiastic about video, particularly entertainment-oriented video, other services left them cold. As a result, cable operators began concentrating their cable system design efforts on efficient, cost-effective video delivery.

Recently, the World Wide Web has generated popular interest in interactive services. Consequently, cable operators are adding high-speed modems and ancillary signals to their offerings. A number of cable operators are also providing telephony services; some major cable operators have merged with telephone companies to merge the two technologies into a unified network.

The late 1980s and early 1990s brought a high level of interest and excitement over the Information Superhighway, also called the National Information Infrastructure (NII). A variety of drivers are energetically pushing the NII. Early efforts came from the telephone industry as it sought to justify upgrading heavily depreciated plant. Since the twisted pair plant did a fine job of delivering plain old telephone service (POTS), another reason for replacing it was needed. The reason had to have a heavy appetite for bandwidth. Digitized voice did not fill the bill since it could be handled over the existing twisted copper pairs. Only computer data and digitized video could provide a justification for declaring the twisted copper pair obsolete. So the telephone systems began a quest for permission to add fiber and digital technology to the rate base. The U.S. Congress consequently decided that a vast government-sponsored program was needed to build the information superhighway and the NII was born.

The cable industry realized that it had a major advantage in this race since it passes 96% of all TV households with a truly wide band delivery means. The cable and telephone industries strove unsuccessfully to convince government that a publicly funded program was not needed. Private industry, they felt could do the job effectively and efficiently.

Since cable television systems must utilize the public right-of-way to install their cables, they, like power, telephone and gas companies, must obtain a franchise from local governmental authorities. This is a non-exclusive franchise. However, experience with multiple cable systems has shown that the economics of the business generally only support one system per community.

Spectrum Reuse

Compared with nearly any other communications need, video is a bandwidth hog. While telephone-quality voice needs only 3 kHz of spectrum and high fidelity sound takes 20 kHz or so (40 kHz for stereo), the current video standard consumes 4.2 MHz. High definition television (HDTV) requires about 30 MHz for each of the red, green and blue signals that make up a color picture. Extensive bandwidth compression techniques reduce the amount of spectrum required by HDTV to the 6 MHz allowed by the Federal Communications Commission (FCC). These signals must then be modulated onto carriers to deliver multiple signals to the consumer's equipment.

For NTSC, each television channel consumes 6 MHz because of vestigial sideband amplitude modulation, VSB-AM. Compared with double sideband amplitude modulation's need for 8.4 MHz, VSB-AM transmits one complete sideband and only a vestige of the other. (At the time the standard was created, electronics consisted of vacuum tubes. It was important to avoid the complexities of single sideband receivers while not consuming the bandwidth required by double sideband transmission.) VSB was an effective compromise for the constraints of the times. The design requirements of practical filters determined the amount of sideband included. The consumer's receiver selects the channel to be watched by tuning a 6 MHz portion of the assigned spectrum. In the terrestrial broadcast environment, channels must be carefully assigned to prevent interference with each other. The result of this process is that most of the terrestrial broadcast television spectrum is vacant. Although better television antennas and television circuits would allow more of the spectrum to be utilized, with 250 million receivers and more than 150 million VCRs in consumers' hands, the changeover process to upgraded systems would be difficult, costly and require approximately 20 years.

The rest of the terrestrial spectrum that is not assigned to broadcast has other important uses. These include aircraft navigation and communications, emergency communications, and commercial and military applications. The terrestrial spectrum is too limited to supply the video needs of the U.S. viewer.

Cable television is made possible by the technology of coaxial cable. Rigid coaxial cable has a solid aluminum outer tube and a center conductor of copper-clad aluminum. Flexible coaxial cable's outer conductor is a combination of metal foil and braided wire, with a copper-clad, steel center conductor. The characteristic impedance of the coaxial cable used in cable television practice is 75 Ω. The well-known principles of transmission line theory apply fully to cable television technology.

The most important characteristic of coaxial cable is its ability to contain a separate frequency spectrum and maintain the properties of that separate spectrum so that it behaves like over-the-air spectrum. This means that a television receiver connected to a cable signal will behave as it does when connected to an antenna. A television set owner can become a cable subscriber without buying additional consumer electronics equipment; much of the cable service can be enjoyed simply by connecting the TV or VCR to the cable system. Other services are enjoyed through adapters provided as part of the cable subscription. The subscriber can also cancel the subscription and not be left with useless hardware. This ease of entry and exit from an optional video service is a fundamental part of cable's appeal to subscribers.

Since the cable spectrum is tightly sealed inside an aluminum environment (the coax cable), a properly installed and maintained cable system can use frequencies assigned for other purposes in the over-the-air environment. This usage takes place without causing interference to these other applications and vice versa. New spectrum is "created" inside the cable by the reuse of spectrum. In some cable systems, dual cables bring two of these sealed spectra into the subscriber's home, with each cable containing different signals.

The principal negative of coaxial cable is its relatively high loss. Coaxial cable signal loss is a function of its diameter, dielectric construction, temperature and operating frequency. A ballpark figure is 1 dB of loss per 100 ft (30.5 m). Half-inch diameter aluminum cable has 1 dB of attenuation per 100 ft (30.5 m) at 181 MHz; at 1 in. (2.5 cm) diameter, the attenuation drops to 0.59 dB per 100 ft (30.5 m). The attenuation of cable varies with the square root of the frequency. Thus, the attenuation at 216 MHz (within TV channel 13) is twice that of 54 MHz (within TV channel 2) since the frequency is four times as great. If channel 2 is attenuated 10 dB in 1,000 ft (.3 km), channel 13 will be attenuated 20 dB.

CABLE NETWORK DESIGN

While current cable practice involves extensive use of fiber in new construction and upgrades, it is important to understand cable techniques used prior to fiber's introduction. This is partly because a significant fraction of cable systems have not yet upgraded to fiber and because these older cable techniques illustrate important cable technical principles.

Since cable television of the 1980s was not a general-purpose communications mechanism, but rather a specialized system for transmitting numerous television channels in a sealed spectrum, the topology or layout of the network was customized for maximum efficiency. The topology that has evolved over the years is called *tree-and-branch architecture*. Many small and intermediate sized systems fit this model. When analyzed, most large systems can be seen as having evolved from this prototype.

There are five major parts to a traditional coaxial cable system:

- the headend
- the trunk cable
- the distribution (or feeder) cable in the neighborhood
- the drop cable to the home and in-house wiring
- the terminal equipment (consumer electronics).

Flexible coaxial cable is used to bring the signal to the terminal equipment in the home. In the simplest cases, the terminal equipment is the television set or VCR. If the TV or VCR does not tune all the channels of interest because it is not cable compatible, a converter is placed between the cable and the TV or VCR tuner.

Broadcast channels 2 through 13 are not in a continuous band. Other radio services occupy the gaps. Cable can reuse these frequencies because its spectrum is self-contained within the coaxial environment. The cable converter has a high quality broadband tuner and output circuitry which puts the desired cable channel on a low band channel not occupied in the local off-the-air spectrum. Typically this is channel 2, 3, 4 or 5. The TV or VCR is tuned to this channel and behaves as a monitor. If programming of interest to the subscriber is scrambled, a descrambler is required; it is usually placed in the converter. Figure 6.13-1 shows the cable frequency plan.

The home is connected to the cable system by the flexible drop cable, typically 150 ft (45.7 m) long. See Figure 6.13-2.

The distribution cable in the neighborhood runs past

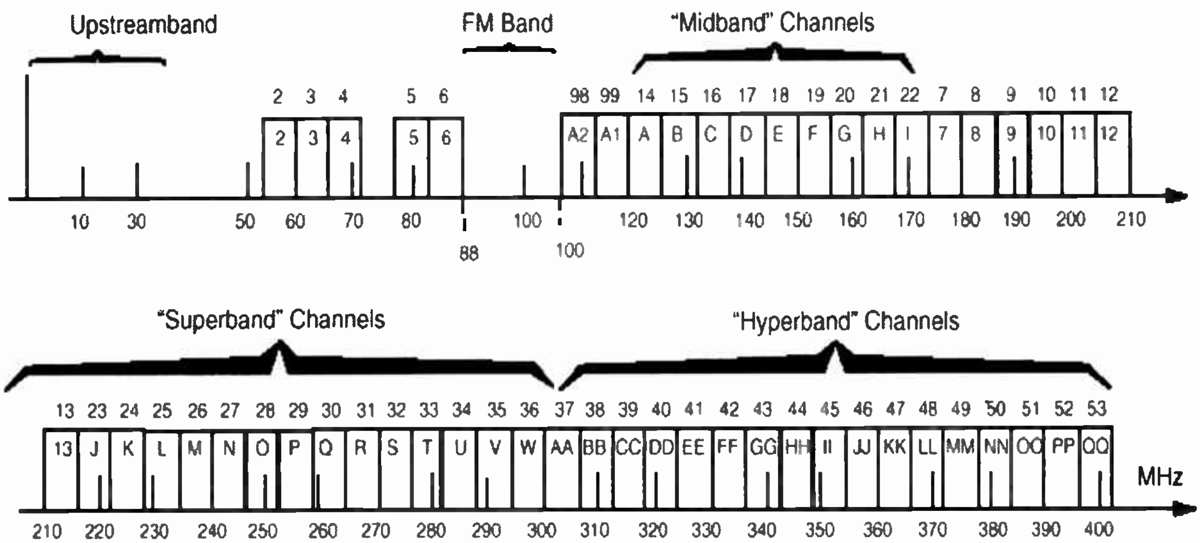


Figure 6.13-1. Frequency plan. Numbers above the rectangles are the Electronics Industry Association designations. Historical designations are inside the rectangles. While the upper frequency shown is 400 MHz, modern cable systems have bandwidths approaching 1 GHz.

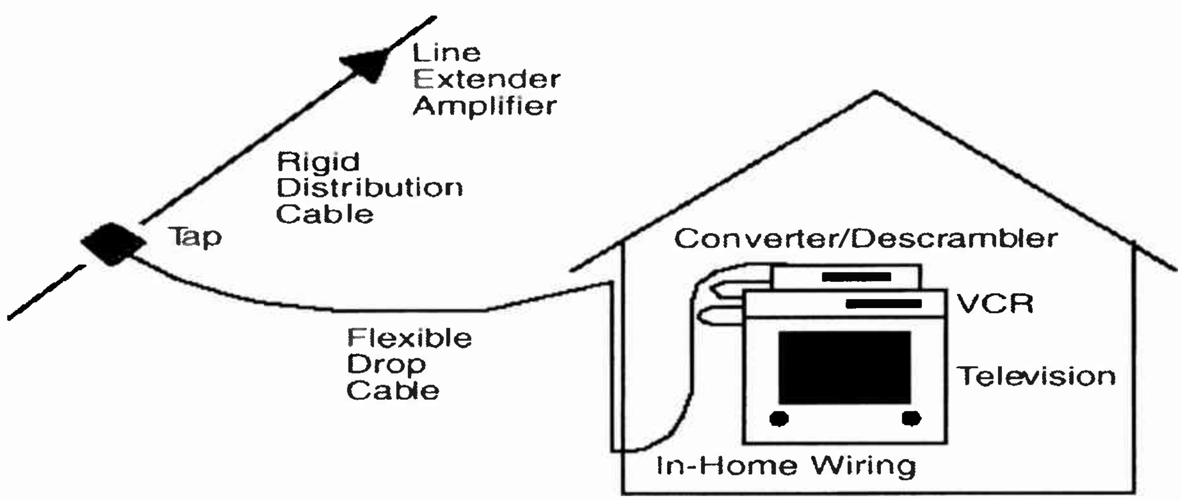


Figure 6.13-2. Terminal equipment and cable drop.

the homes of subscribers. This cable is tapped so that flexible drop cable can be connected to it and routed to the residence. The distribution cable interfaces with the trunk cable through an amplifier called a *bridger amplifier,* which increases the signal level for delivery to multiple homes. One or two specialized amplifiers called *line extenders* are included in each distribution cable. Approximately 40% of the system's cable footage is in the distribution portion of the plant and 45% is in the flexible drops to the home (see Figure 6.13-3.)

The trunk part of the cable system transports the signals to the neighborhood. Its primary goal is to cover distance while preserving the quality of the signal in a cost-effective manner. Current practice uses fiber to bring the signal to the neighborhood. Older cable systems that haven't been rebuilt or upgraded still have coaxial cable in the trunk portion of the cable plant. With coaxial cable, broadband amplifiers are required about every 2,000 ft (.6 km) depending on the bandwidth of the system. The maximum number of amplifiers that can be placed in a run or *cascade* is limited by the build up of noise and distortion. Twenty or 30 amplifiers may be cascaded in relatively high-bandwidth applications. Older cable systems with fewer channels may have as many as 50 or 60 amplifiers in cascade. Approximately 14% of a cable system's footage is in the trunk part of the system.

The headend is the origination point for signals in the cable system. It has parabolic or other appropriately shaped antennas for receiving satellite-delivered program signals, high gain directional antennas for receiving distant TV broadcast signals, directional antennas for receiving local signals, machines for playback of taped programming and commercial insertion and stu-

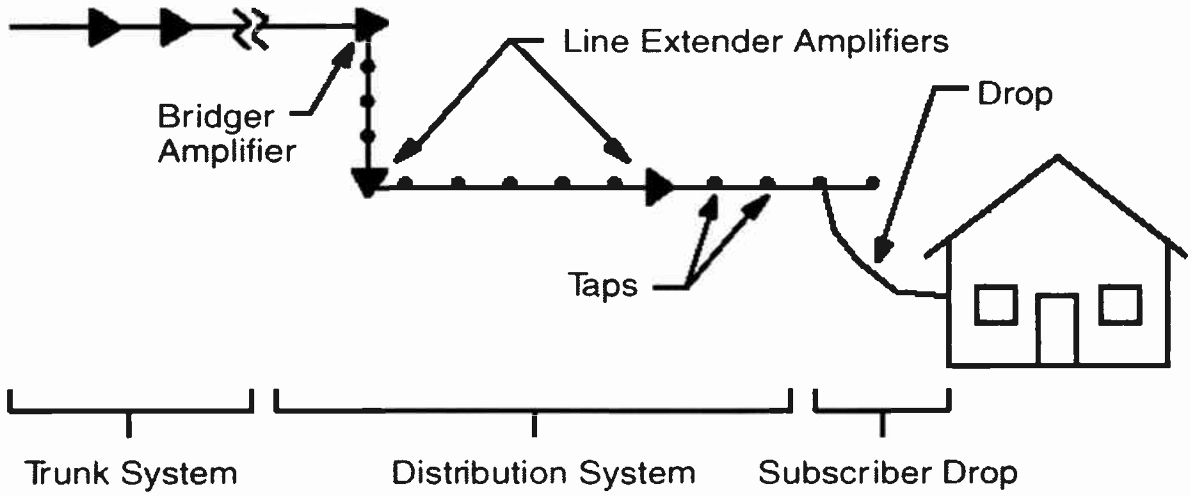


Figure 6.13-3. Distribution plant.

dios for local origination and community access programming (see Figure 6.13-4).

Local origination is programming over which the cable operator has editorial control. It can range from occasional coverage of local events to a collection of programming almost indistinguishable from that of an independent broadcaster. Often mobile coverage of events is provided with microwave links back to the headend or back-feed of the signal up the cable system to the headend.

Community access is channel and/or time availability mandated by the local franchise for programming by community groups. The cable system typically cannot exercise editorial control over quality or content of community access programming.

When the whole picture is assembled, the tree shape of the topology is evident. The trunk and its branches become visible. See Figure 6.13-5.

Signal Quality

The ultimate goal of the cable system is to deliver pictures of adequate quality at an acceptable price while satisfying stockholders, investors and holders of the debt generated by the huge capital expenses of building the cable system plant. This is a difficult balancing act. It would be a simple matter to deliver very high quality video if cost were not a consideration. Experience teaches that subscriber satisfaction is a complex function of a variety of factors led by program quality and variety, reliability of service, video and sound quality and the amount of the subscriber's cable bill.

The principal picture impairments can be divided into two categories—coherent and noncoherent. Coherent impairments result in a recognizable interfering pattern or picture. They tend to be more objectionable than non coherent impairments of equal strength.

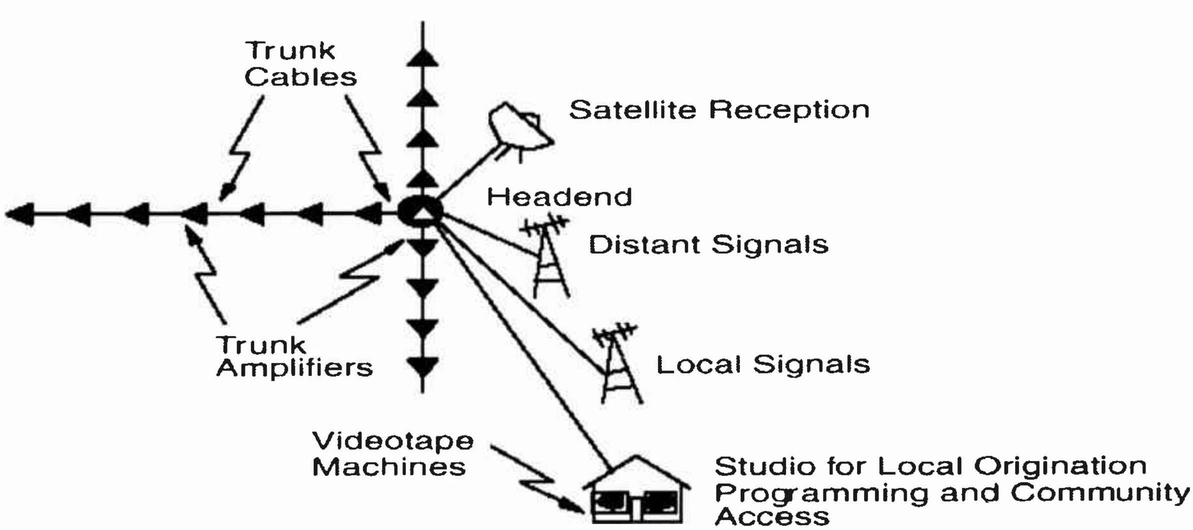


Figure 6.13-4. Cable system headend.

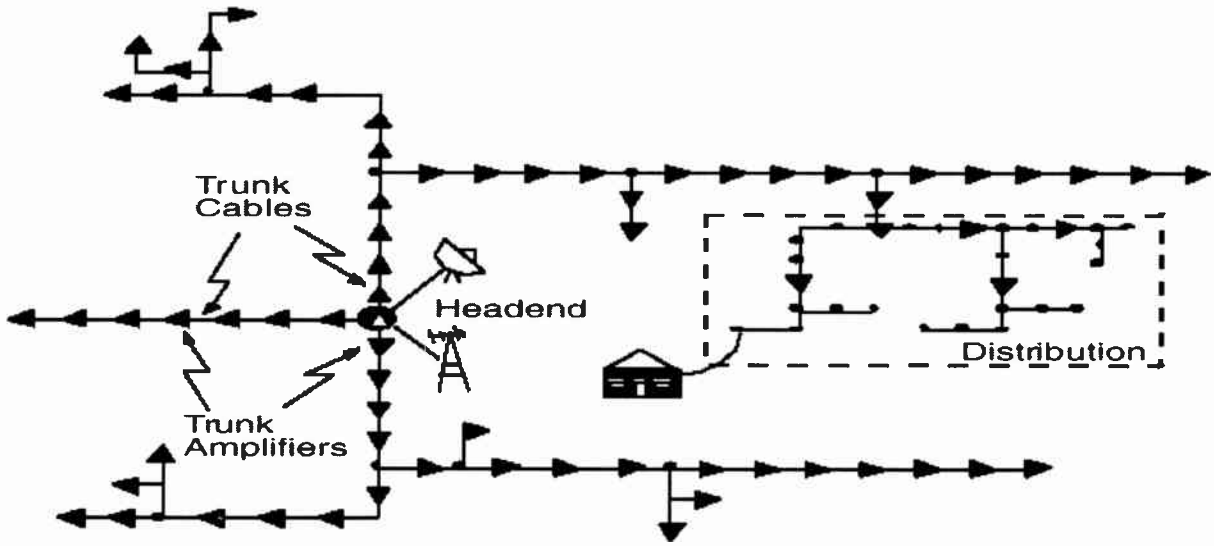


Figure 6.13-5. Tree-and-branch topology.

The principal noncoherent picture impairment is noise. Random noise behavior is a well-understood part of general communications theory. The familiar Boltzmann relationship (noise figure concepts) applies fully to cable television technology. Random noise is the consequence of the statistical nature of the movement of electric charges in conductors. This noise is inescapable and creates a signal of its own. If the intended analog signal is ever allowed to become weak enough to be comparable to the noise signal, it will be polluted by it, yielding a snowy pattern in pictures and a seashore-sounding background to audio.

Noise levels are expressed in cable system practice as ratios of the video carrier to the noise in a television channel. This measure is called the carrier-to-noise ratio (CNR) and is given in decibels (dB). The target value for CNR is usually 45–46 dB or higher. Noise in the picture, called snow, is just visible when CNR is 43–44 dB. Snow is objectionable at CNRs of 40–41 dB.

Coherent interference includes ingress of video signals into the cable system, reflections of the signal from transmission-line impedance discontinuities, cross modulation of video and cross modulation of the carriers in the video signal. This latter phenomenon gives rise to patterns on the screen which are called *beats*. These patterns often look like moving diagonal bars or herringbones.

The evaluation of signal quality takes place on two planes, objective and subjective. In the objective arena, measurements of electrical parameters are used. These measurements are repeatable. Standardized, automated test equipment has been developed and accepted by the video industries. Table 6.13-1 lists the parameters usually considered important and the values of good, current practice. They are described in the remaining text.

The ultimate performance evaluation involves the subjective reaction of viewers. One example of the difficulties experienced is the fact that different frequencies of noise have differing levels of irritation. High frequency noise tends to become unnoticeable while low frequency noise creates large moving blobs which are highly objectionable. Subjective reactions to these phenomena are influenced by such factors as the age, gender, health and attitude of the viewer. The nature of the video program, the characteristics of the viewing equipment and the viewing conditions also impact the result.

Table 6.13-1
Signal quality target values.

Parameter	Symbol	Value
Carrier/Noise (CNR)	C/N	48–50 dB
Composite Second-Order	CSO	−53 dB
Composite Triple Beat	CTB	−53 dB
Signal Level at TV		0–+3 dBmV

Signal processing in the TV receiver changes the impact of signal impairments. Noise in the band of frequencies used to transmit color information is demodulated and converted into lower frequency, more objectionable noise. Noise in the synchronization part of the TV signal can cause the picture to break up entirely, resulting in much greater impairment than the same strength noise confined to other portions of the signal.

In 1959, the Television Allocations Study Organization (TASO) studied the amount of noise, interference and distortion viewers will tolerate in a TV picture. The results were expressed in a five-point scale with grades named excellent, fine, passable, marginal and inferior. These are very old data.

It is important to realize that the demand for signal quality is a function of time. Fifteen to 20 years ago, consumer electronics products were not capable of displaying the full resolution of the NTSC signal.

Gradually, these products improved with high-end models capable of more performance than the NTSC signal can deliver. The Super VHS and Hi-8 video tape systems have greater resolution than broadcast NTSC. As time progresses, the level of performance of consumer electronics will continue to increase. As digital television (DTV) and HDTV are introduced, still more demands will be made on cable system performance. The trend to larger screen sizes also makes video impairments more evident.

Cable System Tradeoffs

The experienced cable system designer of the 1980s learned how to balance noise, nonlinear distortions and cost to find a near optimal balance. The various configurations of then available components were thoroughly explored and the best choices understood. The explosion of technology in the 1990s changed all of this carefully sorted-out technique and dramatically increased the options available to the cable system designer. Construction of optimum designs now requires knowledge, experience, creativity and a complete familiarity with the latest components.

Signals in cable systems are measured in dB relative to 1 mV across 75 Ω. This measure is called dBmV. Applying the Boltzmann noise equation to 75 Ω cable systems yields an open-circuit voltage of 2.2 microvolts in 4 MHz at room temperature. When terminated in a matched load, the result is 1.1 microvolts. Expressed in dBmV, the minimum room-temperature noise in a perfect cable system is −59.17 dBmV.

Starting at the home, the objective is to deliver at least 0 dBmV, but no more than 10 dBmV to the terminal on the television receiver. Lower numbers produce snowy pictures, and higher numbers overload the television receiver's tuner, resulting in cross modulation of the channels. If a converter or descrambler is used, its noise figure must be taken into account. There are two reasons for staying toward the low side of the signal range: cost and the minimization of interference in the event of a signal leak caused by a faulty connector, damaged piece of cable or defects or inadequate shielding in the television receiver. Low signal levels may cause poor pictures for the subscriber who insists on unauthorized splitting in the home to serve multiple receivers. Working our way back up the plant, we need a signal level of 10 dBmV to 15 dBmV at the tap to compensate for losses in the drop cable.

The design objectives of the distribution part of the cable system involve an adequate level of power not only to support the attenuation characteristics of the cable but to allow energy to be diverted to subscribers' premises. Energy diverted to the subscriber is lost from the distribution cable. This loss is called *flat loss* because it is independent of frequency. Loss in the cable itself is a square-root function of frequency and is therefore contrasted to flat loss. Because of flat losses, relatively high power levels are required in the distribution part of the plant, typically 48 dBmV at the input to the distribution plant. These levels force the amplifiers in the distribution part of the plant to reach into regions of their transfer characteristics which are slightly nonlinear. As a result, only one or two amplifiers, called line extenders, can be cascaded in the distribution part of the plant. These amplifiers are spaced 300–900 ft (91.4 m–274.3 m) apart depending on the number of taps required by the density of homes.

Because the distribution part of the plant is operated at higher power levels, nonlinear effects become important. The television signal has three principal carriers: the video carrier, the audio carrier and the color subcarrier. These concentrations of energy in the frequency domain give rise to a wide range of *beats* when passed through nonlinearities. To minimize these effects, the audio carrier is attenuated about 15 dB below the video carrier.

When cable systems only carried the 12 VHF channels, second-order distortions created spectrum products which fell out of the frequency band of interest. As channels were added to fill the spectrum from 54 MHz to as much as 750 MHz (1 GHz in a couple of systems), second-order effects were minimized through the use of balanced, push-pull output circuits in amplifiers. The third-order component of the transfer characteristic dominates in many of these designs. The total effect of all the carriers beating against each other gives rise to an interference called CTB. CTB is measured with a standard procedure involving 35-channel carriers. In a 35-channel cable system, about 10,000 beat products are created. Channel 11 suffers the most with 350 of these products falling in its video passband. Third-order distortions increase about 6 dB for each doubling of the number of amplifiers in cascade. A 1 dB reduction in amplifier output level will generally improve CTB by 2 dB. If these products build to visible levels, diagonal lines will be seen moving through the picture. When these components fall in the part of the spectrum that conveys color information, spurious rainbows appear.

If we assign a design level of noise and nonlinear distortion at the subscriber's television receiver that is below the threshold of visibility, we can create a budget of noise and distortion to be "spent" in the various parts of the system design. The distribution part of the system has relatively high powers and eats up most of the budget for nonlinear distortions. On the other hand, little of the noise budget is consumed. It can be allocated to the trunk part of the system, which brings the signals into the neighborhood.

The design objective of the trunk part of the cable system is to move the signal over substantial distances with minimal degradation. Because distances are significant, fiber or lower-loss cables are used. One-in. (2.5 cm) and 0.75 in. (1.9 cm) diameter cable is common in the trunk while 0.5-in. (1.3 cm) cable is found in the distribution. Signal levels in the trunk at an amplifier's output are 30–32 dBmV depending on the equipment used.

It has been determined through analysis and confirmed through experience that optimum noise performance is obtained when the signal is not allowed to be attenuated more than about 20–22 dB before being amplified again. Amplifiers are said to be *spaced* by 20 dB. The actual distance in feet is a function of maximum frequency carried and the cable's attenuation characteristic. High-bandwidth cable systems have their amplifiers fewer feet apart than older systems with fewer channels. Since attenuation varies with frequency, the spectrum in coaxial cable develops a *slope*. This is partially compensated with relatively simple equalization networks in the amplifier housings.

The attenuation of the cable is a function of temperature and aging of components. These amplifiers use a pilot signal to control automatic gain control (AGC) circuits. A second pilot signal at a substantially different frequency than the first allows the slope of the attenuation characteristic to be monitored and compensation to be introduced with automatic slope control (ASC) circuits. Thus, long cascades of amplifiers can, once properly set up, maintain their performance over practical ranges of temperature and component aging.

Since the signal is not repeatedly tapped off in the trunk part of the system, high power levels are not required to feed splitting losses. As a result, signal levels are lower than in the distribution portion of the plant. Typical levels are about 30 dBmV. For the most part, the amplifiers of the trunk are operated within their linear regions. The principal challenge of trunk design is keeping noise under control. Each doubling of the number of amplifiers in the cascade results in a 3 dB decrease in the CNR at the end of the cascade and a 6 dB increase in the amount of CTB.

If the noise at the end of the cascade is unacceptable, the choices are to employ lower noise amplifiers, shorter cascades or a different technology such as microwave links or fiber optic links.

Cable Economics

The most recent figures compiled (for 1997) indicate that the U.S. cable industry has 11,517 cable systems. These range in size from a few hundred customers to the largest system, which serves about a million subscribers in New York City and is operated by Time Warner. The cable industry employs about 125,000 people and has annual revenues of $33 billion ($25 billion from subscriptions and about $8 billion from advertising).

The economics of cable television is governed by a number of factors. Included in these are the cost of construction, the cost of programming, the willingness of subscribers to pay for the service and the operating costs of the system.

It is difficult to define a typical cable system because of the variety of local factors. However, a rough approximation of such an entity is a 54-channel two-way capable, one-way activated system with 50,000 subscribers. Population density is about 75 homes per mile of cable plant. Penetration of subscribers is about 67% of homes passed. The current cost to build such a system is between $600 and $1,200 per subscriber, depending on local geography, economics and construction rules. The cost of cable system construction varies greatly depending on the circumstances of the location. A major variable is whether the cable system plant is underground or on poles; in the United States, only about 20% of the plant is underground.

Figure 6.13-6 shows an approximate distribution of cable construction costs. The labor to install each component is included with the cost of that component. Hardware includes the mechanism for supporting the cable. In aerial construction, a strong, heavy wire called *strand* is mounted on the poles. Strand also supports amplifiers and passive components. The passive devices split the signal and tap into the cable for distribution to the home. A thinner wire, called *lashing*, is spiral wrapped around the cable and strand to support the cable. Electronics includes amplifiers and microwave links.

The market value of cable systems recently sold varies from $1,800 to a high of almost $2,500 per subscriber. These values are determined by both the current cash flow and the potential for increased penetration and new revenue streams. The capital-intensive cable business is currently experiencing dramatic increases in programming costs.

According to statistics gathered in 1997 by the NCTA, the cable industry trade organization subscribers paid about $27 per month for basic service and $8.29 per month for each optional pay service such as HBO. While not every basic subscriber takes an optional pay service, some take more than one. Average per month subscription revenue is about $37.75.

TECHNICAL DETAIL

This section concentrates on issues relating to the technical performance of cable systems. Some generaliza-

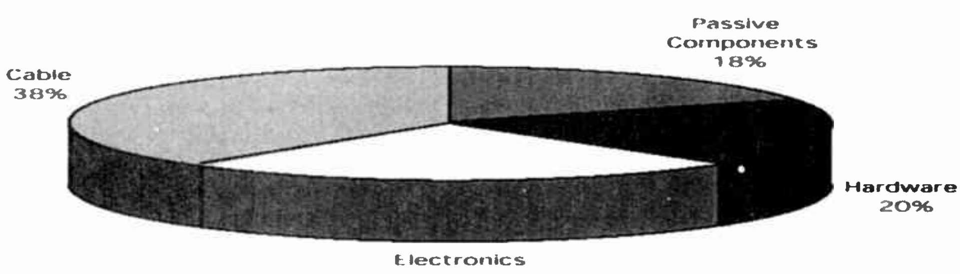


Figure 6.13-6. Distribution cable plant costs.

tions have been made in order to group explanations. This section is intended to serve as a briefing, so selective tradeoffs were made on the amount of detail given. There are always exceptions, for cable systems do not neatly fall into clear types.

System Configurations and Trends

A cable system's configuration consists of:

- The headend (the signal reception, origination, and modulation point)
- Main, coaxial trunk (or tree) cable, which runs through central streets in communities
- Coaxial distribution (branch) cable to the customer's neighborhood, including distribution taps
- Subscriber drops to the house
- Subscriber terminal equipment (television sets, converter/descramblers, VCRs).

Distribution plant is sometimes called *feeder* plant. Programming comes to the headend by satellite signals, off-air signals from broadcast stations and signals imported via terrestrial microwave. Signals originating from the headend are from a co-located studio facility, VCRs, character generators or commercial insertion equipment.

Plant mileage is calculated using the combined miles (kilometers) of strand that support the coaxial cables in the air and the footage of trenches where cables are installed in the ground. There are about a million miles (1.6 km) of plant in the 11,517 U.S. cable systems in existence in 1997.

Extension cables, or *drops*, interconnect main coaxial plant lines to customers' homes. They are not included in plant mileage. Drop cables are smaller in diameter than mainline coaxial cable plant. They interconnect between a *power splitter*, called a *multitap directional coupler*, and the customer's interface, usually a television set. The tap is located in the utility easement. In an average cable system there will be 6,625 ft (2 km) of drop cable per plant mile and 7,040 ft (2.1 km) of hard-line coaxial cable. Other interesting statistics include an average of 75 homes per mile, 67% penetration, drop length of 125 ft (38.1 m) and 1:3 trunk footage-to-feeder footage ratio. To put it another way, more than 50% of the total plant is drop cable and associated F-connectors, which are the connectors used to connect coaxial cable to equipment. About half of service calls are related to problems with the drop portion of the plant. About one-third of the drop-related service calls are caused by problems at the F-connectors. Cable systems replace approximately 30% of drops annually.

Industry Trends

ATV will put pressure on operators to expand plant bandwidth.

The last few years have brought exciting trends employing new technologies. Fiber is now installed to upgrade older systems and as part of rebuilds and new builds. The old trunk system of long cascades of amplifiers is now considered obsolete. Work on new amplifier technologies will allow a realization of cable's inherent bandwidth, which exceeds 1 GHz.

A bandwidth of 1 GHz contains 160 slots of 6 MHz. These can be allocated to NTSC, DTV and to new services. The most exciting potential lies with utilizing video compression technology to squeeze 6–12 NTSC-like quality signals in a 6 MHz slot. This opens the door for multiple hundreds of channels. Near video on demand becomes practical (the most popular movies could be repeated every few minutes to minimize the wait time before a movie starts). The average wait time could be made shorter than the trip to the video store, the subscriber would not have to make a second trip to return the movie and the movie would always be in. A microprocessor can keep track of which channel to return to should the subscriber wish to take a break. It is possible to design systems that behave as if they are switched even though they remain more like a traditional cable tree-and-branch structure.

Channelization

There are three channelization plans to standardize the frequencies of channels. The standard assignment plan has evolved from the frequency assignments that the FCC issued to VHF television broadcast stations.

The second channelization plan is achieved by phase locking the television channel carriers. The incrementally related carriers (IRC) plan was developed to minimize the effects of third-order distortions generated by repeated amplification of the television signals as they pass through the cable plant. As channel capacities increased beyond 36 channels, composite, third-order distortions became the limiting distortion.

The third channelization type is the harmonically related carriers (HRC) plan. It differs from the standard and IRC plan by lowering carrier frequencies by 1.25 MHz. With HRC, carriers are phase locked and fall on integer multiples of 6 MHz starting with channel 2 at 54 MHz. This plan was created to further reduce the visible impact of amplifier distortions.

The channelization plans were designed to reduce the visibility of distortion products by making their frequencies synchronous with the interfered carrier. Since carriers present in the downstream signal path add to the distortions, cable systems carry nonvideo carriers at a level that is 13–17 dB below the video carrier's amplitude. This drastically reduces distortion contributions. With broadcast TV channels carried on cable, special processing equipment is used to reduce the aural carrier. This amplitude reduction does not significantly affect the audio signal-to-noise ratio (SNR) quality of monaural television sound. However, this lower level created SNR problems at the end of the cable system for higher bandwidth signals such as FM stereo. When stereo television audio was developed, careful attention was focused on encoding techniques that would promote SNR immunity of the difference channel.

FM radio services are carried at an amplitude that is 15–17 dB below channel 6's video carrier level. The services are carried on cable in the FM band slot of 88–108 MHz. In an IRC channel plan, channel 6's aural carrier falls at 89.75 MHz, which reduces the available FM band to 90–108 MHz.

Low-speed data carriers are transmitted in the FM band or in the guard band between channels 4 and 5 in a standard frequency plan. The amplitude of these carriers is at least 15 dB below the closest video carrier level.

Signal Transportation Systems

Transportation systems were developed to deliver high quality signals from the central headend point to remote headend or hub locations where cable signals are injected into cable trunking systems. The increase in channel capacity and subsequent need to decrease amplifier cascades directly affected the development of transportation systems. Urban franchises cover large areas yet wish to transmit downstream signals from a common headend point. Other motivating factors include increased local programming originating from the cable system's studio, the requirement to deliver city government programming originating from municipal locations, a new business opportunity for the insertion of local commercials to satellite-delivered services and the development of PPV programming.

Several transportation methods have become popular over the last 10 years: amplitude modulated microwave link (AML), frequency modulated microwave link (FML), frequency modulated coaxial link (FMCL), amplitude modulated coaxial supertrunks and fiber interconnects.

AML and FML Microwave Links

AML allows the delivery of the entire downstream cable spectrum through the air to reception points located 8–10 miles (12.9–16.1 km) away from the microwave transmit site. An AML microwave transmitter provides adequate power to deliver signals to eight reception hubs. The advantage of AML is that the receiver simply performs a wide band, block down conversion. The entire downstream frequency spectrum (50–400+ MHz) is reproduced at an appropriate level for direct insertion into the cable plant, without the need for individual channel frequency shifting or processing. The only equipment required at an AML reception site is the microwave receive antenna and a microwave receiver. These can be mounted to a small tower or telephone pole.

AML up converts each television channel and combines it at its respective microwave frequency, with one frequency per channel, using a complex wave guide system. In some cases, each channel added lowers the maximum power per channel. This may either shorten path lengths or increase the probability of signal degradation during rain fades.

One of the main uses of FML is to cover distances not feasible with AML or where video SNR in excess of 56 dB is required. FML is a single-channel, frequency-modulated microwave transmission system. FML occupies significantly more bandwidth than amplitude modulated systems, from 12.5–25 MHz per channel, depending on path length. Fewer channels are available for use. Since this transmission system delivers FM signals, demodulation and VSB-AM and frequency translation must occur before insertion into the cable plant. The FML system is often used to deliver specialized programming (local origination) to a remote headend.

Amplitude Modulated Coaxial Supertrunk

The amplitude modulated supertrunk transportation system can be as simple as a conventional coaxial trunk transporting VSB-AM signals between two headends. Or the supertrunk can link a primary headend directly to the beginning of the normal cable distribution trunk. The television channels can be combined at their final frequency assignments for direct insertion into the distribution trunk. Or they can be grouped to fit into a reduced coaxial bandwidth (5–108 MHz) for transportation on a special service coaxial trunk. An example of this trunk is a mid-split system, which is a bi-directional trunk using 158–330 MHz downstream and 5–108 MHz upstream.

Another variation of AM supertrunk is split-band trunking. The spectrum from 50–450 MHz is split into two groups of 30 television channels each and inserted on two side-by-side trunks that connect two sites. The signals are equally loaded onto two trunk cables to improve television signal quality. This is achieved because the distortion produced is directly related to the number of channels carried on each trunk cable.

Whenever signals are not transported on their final frequency assignment, frequency translation devices called *channel processors* are used to frequency shift the television signals. AM supertrunks can be similar to cable distribution trunks. The difference is that the AM supertrunk uses high technology trunk amplifiers, such as *feedforward*. These transport with significantly less distortions and better CNR.

Fiber Interconnect

Fiber optic transportation systems are now used. Currently, analog video fiber technology is preferred for this application, although occasionally cable systems use digital video. The focus of current research is to optimize fiber optic transmitters, amplifiers and receivers.

The initial application of this technology took place in Orlando, FL, in 1988 as a backup to microwave links. Heavy rain caused these links to fade or suffer complete blockage. Automatic equipment switches to the fiber link. The AM fiber technology is now commonly used instead of microwave in many applications.

Digital fiber links are used when video signals are partially transported in a digital common carrier network. The video interfaces used operate at a DS-3 rate of 45 Mbps, which routes through common carrier points of presence and switching networks.

Frequency Band Usage

There are several ways to increase channel capacity. If the actual cable is in good condition, channel capacity is upgraded by modifying or replacing the trunk and distribution amplifiers. If the cable has seriously deteriorated, the cable plant is completely rebuilt.

Upgrades (Retrofitting) and Rebuilds

An upgrade is defined as a plant rehabilitation process that results in the exchange or modification of amplifiers and passive devices (such as line splitters, directional couplers and customer multitaps). Nearly all upgrades involve adding (or increasing the amount of) fiber from the headend to the neighborhood. A simple upgrade requires new amplifier circuit units called *hybrids*. A full upgrade replaces all devices in the system. In an upgrade project, most of the feeder cable is retained. Goals of an upgrade project include increasing the plant's channel capacity and system expansion to outlying geographic areas. Upgrades are often the least expensive solution to providing expanded service.

In a rebuild, the outside plant is replaced. Customer drops are replaced on an as-needed basis. The strand that supported the old cable is occasionally retained. A rebuild is the most expensive solution to providing upgraded service. It requires a minimum of system downtime, since both old and new plant are active for a period of time. This allows the customer's drop to be switched directly from the old system to the new.

Once the plant has been rebuilt or upgraded, customers are provided newer converters with additional capabilities. The displaced units are moved to other systems or used for the basic tier of service.

System Distortion and System Maintenance

Constraints on the design and implementation of cable systems are imposed by each device used to transport or otherwise process the television signal. Each active device adds small distortions and noise to the signal. Even passive devices contribute noise. The distortions and noise compound so that with each additional device the signal becomes less perfect.

Any nonlinear device, even bimetallic junctions, cause distortions. The primary contributors are the slight nonlinearities of amplifiers. Because the amplifiers are connected in cascade, the slight damage to the signal accumulates.

Noise in any electronic system can come from many sources. The major source is the random thermal movement of electrons in resistive components. For a cable system at 20°C or 68°F, the thermal noise voltage in a single channel's bandwidth will be 1.1 microvolts or −59.1 dBmV. This is the minimum noise level or noise floor. Noise contributions from amplifiers add on a power basis, with the noise level increasing 3 dB for each doubling of the number of identical amplifiers in cascade. Eventually, the noise will increase to objectionable levels. The difference between the radio frequency (RF) peak level and the noise level is measured to quantify the degree of interference of the noise power. The power levels in watts are compared as a ratio. This is called the signal-to-noise ratio (S/N). In a cable system, the apparent effect of noise is its interference with the video portion of the TV channel. This level is compared to the video carrier and called the CNR.

As the CNR value decreases, the interference of noise with the signal becomes visible as a random fuzziness or *snow*, that can overwhelm the picture resolution and contrast. The point where the picture becomes objectionably noisy to viewers is at a CNR of 40 dB. In well-designed systems, the CNR is maintained at 46 dB, preferably at 48–50 dB. While an increase in signal level would improve the CNR, unfortunately, there can be no level increase without increases in distortions.

The distortion products of solid-state devices used in cable amplifiers are a function of the output levels and bandwidths. The higher the signal level, the greater the distortion products produced. Modern amplifiers use balanced configurations which almost completely cancel the distortion caused by the squared term of the amplifier's transfer characteristic. The dominant remaining distortions are called triple beats and are caused by the cubed term. Because distortion products add on a voltage basis, the CTB-to-carrier ratio changes by 6 dB for each doubling of the number of amplifiers in cascade, whereas the CNR decreases by 3 dB for each doubling.

The operating parameters of amplifiers determine the number that can be cascaded and, hence, the distance that can be covered. This is critical to the design philosophy of coaxial trunks in cable systems. While fiber is now used in trunk construction and rebuilds, the following discussion demonstrates important cable principles nonetheless. It also helps explain the importance of fiber in overcoming coaxial cable's limitations.

A cable system's operating limits are defined in terms of its noise floor and distortion ceiling. The noise floor and distortion ceiling are a function of the number of amplifiers in cascade for a system at 300 MHz. A cascade of 46 trunk amplifiers is possible while realizing a 46 dB CNR and a 53 dB CTB. However, other operating realities dictate that substantially more headroom for both distortion and noise be incorporated into the design. The other factors to consider include: change in cable attenuation and noise with temperature, AGC and ASC tolerances, system frequency response, accuracy of field test equipment, designs anticipating DTV and maintenance probabilities. In the example cited, allowing 1 dB of AGC/ASC change, a 3 dB peak-to-valley and 2 dB of test equipment uncertainty results in a 6 dB tolerance. Limiting distortion, CTB, implies that the cascade should only be half the length predicted in the chart, or 23 amplifiers.

The foregoing discussion is applicable only to the trunk portion of the system. As signal levels are increased in the distribution sections, additional allowances must be made in the system design. As a rule of thumb, CNR is determined primarily by the conditions of trunk operation, and signal-to-distortion ratio (SDR) primarily by the conditions of distribution operation.

Two other factors limit the geography of a cable system. Cable attenuation rises with increasing frequency. More equal gain amplifiers are required to transmit the signal a given distance. But noise limits the maximum number of amplifiers used. The second factor is that amplifier distortion is a function of channel loading: The more channels carried, the greater the distortions.

To obtain optimum cascade length, AGC/ASC tolerance, accurate alignment, calibrated test equipment and well-founded system maintenance programs are of paramount importance. Maintenance programs are designed to ensure that system alignment is kept within acceptable limits. Where trunk lines carry signals through cascades of up to 40 trunk amplifiers, it is crucial that each amplifier have a flat amplitude-versus-frequency response. The additive effects of even minimal amplitude response variations in each amplifier create significant system flatness problems at the end of long cascades.

Maintaining Amplitude Versus Frequency Response

A maintenance program objective should be to achieve a system amplitude versus frequency response of less than $[(N/10) + 1]$ dB peak-to-valley. This will minimize the degradation of CNR which occurs in channels that fall in the valleys of the system response. Here N is the number of trunk amplifiers in the cascade. It is imperative that the optimum response is maintained at each amplifier. Recommended practice calls for no more than 3 dB difference in adjacent channel video carrier amplitudes provided to the customer. All channels must fall within a 12 dB overall passband response window.

The common method of evaluating a system's frequency response is a *sweep generator* that injects a rapidly swept carrier over the system's passband. The sweep generator is set to sweep from 50 MHz to the upper system passband frequency with a duration time as short as 1 millisecond. The field sweep receiver is then synchronized to the generator. The portable receiver provides a display of the system's response at each amplifier as the maintenance technician progresses through the trunk cascade.

Another tool for checking system flatness is the *spectrum analyzer*. To determine the overall system response, the individual video carrier amplitude of each television channel is measured. Because the carrier amplitudes are usually adjusted for a sloped amplifier output to minimize distortion products, more interpretation is needed with a spectrum analyzer. The sweep system allows the amplifiers to be adjusted for the flattest response.

Excessive response variations can cause additional distortions since some carriers will now exceed the amplitude at which the system's amplifiers were designed to operate. To keep these response variations from becoming excessive, amplifier manufacturers make a response control device known as a *mop-up circuit*. It is installed at periodic locations in trunk amplifiers throughout the cascade. These mop-up circuits are tunable filters adjusted to remove small peaks (less than 1.5 dB) caused by the amplifiers or cable. Incorrect use of these devices to solve defective equipment problems (bad cable sections, splices or line passives) can cause impairment to a video channel by changing its in-band frequency response or its chrominance-to-luminance delay characteristics.

Group Delay through the Cable Plant

Trunk amplifiers with bidirectional capability exhibit group delay as a result of the band-splitting diplex filters. The visible effect of the filtering is a loss of resolution in the picture. The diplex filters, which are high-pass and low-pass filters with a 40 MHz crossover frequency, are part of the trunk amplifier's circuitry.

Channels 2, 3 and 4 suffer from the repeated effects of the filtering. Other locations where filtering occurs are apartment complexes, hotels/motels and hospitals, where channels are deleted from the spectrum by special band stop filters. Locally originated channels are inserted in the deleted portions of the spectrum.

Filtering occurs at the headend or hub in connection with channel processing equipment or the channel modulators. The effects of these filters are taken into account when the headend equipment is designed. With only one of these devices per channel, the delay effects of this equipment rarely create problems. However, with some configurations of signal transportation systems, additional single-channel or multiple-channel filtering may take place and cause delays at hubs.

System Reflections

Signal reflections occur throughout the cable plant and are called *microreflections*. They are caused by the individual slight errors in impedance match. The severity of the mismatch is measured by the magnitude of the return-loss ratio. The larger the return loss, the better. Perfection is infinite. Mismatches include connectors, splices and even damage to the cable itself. As an example, if a splice installed in a trunk line approximately 150 ft (45.7 m) past an amplifier. The splice, which may only have a 12 dB return loss, reflects signals back upstream that have only been attenuated by 13.5 dB (1.5 dB in cable attenuation plus the return loss of the splice). The reflected signals then arrive back at the output of the amplifier attenuated by a total of 15 dB (1.5 dB additional cable loss for the upstream trip plus the previous 13.5 dB). The signals are now reflected by the amplifier's output mismatch; a return loss of 16 dB is common. At this point, the reflected signal has an amplitude that is 31 dB below the primary signal and delayed by the round-trip propagation through 300 ft (91.4 m) of cable, which takes about 350 nanoseconds. The signal is horizontally delayed approximately 1/7 in. (3.6 mm) on a 27-in. (68.6 cm) television set. This is not enough to become a visible ghost or second image. However, depending on the relative phases of the RF carriers of the primary and reflected signal, the visual effect may be enough to cause a softening of a previously well-defined luminance level transition. With repeated mismatches, the crispness of the pictures may be noticeably reduced. This softening effect is easily seen on displays of character generated text pages.

Phase Noise

Phase noise is added to the original signal through modulation and frequency conversion processes. A significant amount of phase noise must be added to the video carrier before generated impairments become perceptible. Narrow band phase noise (measured 20 kHz from the video carrier) in a TV channel produces variations in the luminance and chrominance levels that appear as an extremely grainy pattern within the picture. The perceptibility level of phase noise on the video carrier is 53 dB below the carrier at 20 kHz. If the frequency conversion or modulation processes are operating close to specification, phase noise impairments should not be perceptible on the customer's TV unless the converter/descrambler is malfunctioning or is of poor quality.

Amplifier Distortions and Their Effects

Amplifier technology based on feedforward and power-doubling techniques increase power levels with fewer distortions. However, additional sources of minutely delayed signals have been created. The signal delays produced in these amplifiers have similar end results in picture degradation as the delayed signals generated by reflected signals in the cable plant. But they are caused by a different mechanism. These amplifiers use parallel amplification technology; the signals are split, separately amplified and then recombined.

With a feedforward amplifier, the signals are purposely processed with delay lines. If the propagation time is not identical through each of the amplifiers' parallel circuits, signals will be recombined that are delayed by different amounts of time. In most circumstances, the amount of differential delay is small and will not produce a visible ghost, but it may cause loss of picture crispness. Since the hybrids used in these amplifiers are normally provided in matched pairs or in a single hybrid package, these delays are only a problem when the hybrids are not replaced as a matched set.

In systems that carry more than 30 channels, CTB is the limiting distortion. However, cross modulation (X-MOD) distortion, which is often the limiting factor in systems with less than 30 channels, can reappear as the controlling factor in dictating system design. The HRC and IRC channelization plans discussed previously were developed to minimize the visible degradation in picture quality that is caused by CTB.

X-MOD is one of the easiest distortions to identify visually. Moderate cross-modulation appears as horizontal and vertical synchronizing bars that move across the screen. In severe cases, the video of multiple channels is visible in the background.

Moderate CTB is the most misleading distortion since it appears as slightly noisy pictures. Most technicians conclude that there are low signal levels and CNR problems. CTB becomes visible as amplifier operating levels exceed the design parameters. Once CTB reaches a severe stage, it becomes more readily identifiable because it causes considerable streaking in the picture.

CSO can become a limiting factor in systems that carry 60 or more channels and use the HRC or IRC channelization plans. This distortion appears as a fuzzy herringbone pattern on the television screen. The CSO beats fall approximately 0.75 MHz and 1.25 MHz above the video carrier in a television channel. An IRC channelization will frequency-lock these beats together while increasing their amplitude relative to the carrier level.

Hum modulation caused by the 60-Hz amplifier powering is identified by its characteristic horizontal bar that rolls through the picture. If the hum modulation is caused by the lack of ripple filtering in the amplifier power supply, it will appear as two equally spaced horizontal bars that roll through the picture.

Frequency Bands Affected by RF Interference

Discrete beat products can be difficult to identify by the displayed picture impairment. Radio frequency interference that leaks into the cable system from nearby RF transmitters causes spurious carriers to fall in the cable spectrum. Common sources of signal leakage are cracked cables and poor quality connections. When either of these situations happens, strong off-air television and FM radio broadcast signals interfere.

If television stations are carried at the same frequency on cable as broadcast and the headend channel processing equipment is phase locked to the off-air signal, the effects of this interference will be ghosting. The ghost appears before (to the left of) the cable signal since propagation time through the air is less than through cable. If the signals are not phase locked together, lines and beats appear in the picture.

Under HRC or IRC, the headend channel processing and modulating equipment are already locked to a reference oscillator. It is not possible to lock the unit to both the cable channelization plan and the off-air signal. In the IRC plan, it may be desirable to unlock a channel from the reference source and lock it to the off-air station. However, the group phase locking advantage is lost. Several dB of distortion resistance on that channel is sacrificed. With HRC systems, video carrier frequencies fall 1.25 MHz lower than their off-air channel's counterpart. It is impossible to carry any channel on the off-air assignment unless the upper adjacent channel slot is not used. With the exception of the new 550 MHz cable systems, the interference from these sources is limited to the VHF channels 2–13 and the FM broadcast spectrum of 88–108 MHz.

Often there is interference from off-air signals due to consumer electronics hardware design. If the internal shielding of the equipment is inadequate, the internal circuits will directly pick up the signal. This phenomenon is called *direct pick-up interference* (DPU) and is the original motivation for cable converters. Those early set-top boxes tuned no more channels than the TV set, but they protected against DPU by incorporating superior shielding and connecting to the TV set through a channel not occupied off-air.

DPU can be misleading. When the subscriber switches to an antenna, he might receive better pictures than from the cable connection. He concludes that his TV receiver is operating correctly and the cable system is faulty. The only convincing argument is a demonstration with a receiver that does not suffer from DPU. Viacom Cable has measured off-air field intensities of eight volts per meter. The German specification for immunity to DPU is four volts per meter. The United States has no such specification. However, TV receivers sold in the U.S. are built to comply with the Canadian specification of one-tenth of a volt per meter. This is inadequate. VCR tuners are generally inferior to TV tuners because the VCR market is even more price competitive. The Electronic Industries Association (EIA) and NCTA Joint Engineering Committee (JEC) are studying this issue under their work on IS-23, the Interim Standard on the RF cable interface.

The second most likely sources of radio frequency interference are created by business band radios, paging systems and amateur radio operators. These signals leak into the cable system and interfere with cable channels 18 through 22 and channels 23 and 24 (145–175 and 220–225 MHz). It is easy to determine that these signals are caused by an RF transmitter because of the duration of the interference and, sometimes, by hearing the broadcast audio. Since the signals are broadcast intermittently, it is almost impossible to determine the exact location(s) of ingress. Cable systems that operate above 450 MHz may find severe forms of interference. They are subjected to high-power UHF television stations, mobile radio units and repeaters as well as a group of amateur radio operators signals in the top 10–12 channels. The extreme variation of shortwave signals in time and intensity makes locating the point(s) of infiltration of these signals difficult.

The upstream 5–40 MHz spectrum is a problem for operators who have two-way cable systems. There are many sources of interference and these signals accumulate upstream. In two-way plant, a single leak in the system can make that portion of the upstream spectrum unusable throughout the entire plant; in the downstream spectrum, a leak may only affect a single customer's reception.

Signal Security Systems

Means of securing services from unauthorized viewership of individual channels range from simple filtering schemes to remote controlled converter/descramblers. The filtering method is the commonly used method of signal security and is the least expensive.

Trapping Systems

There are two types of filtering or trapping schemes: *positive trapping* and *negative trapping*. In the positive trapping method, an interfering jamming carrier(s) is inserted into the video channel at the headend. If the customer subscribes to the secured service, a positive trap is installed at the customer's house to remove the interfering carrier. The positive trapping scheme is the least expensive means of securing a channel where less than half the customers subscribe.

A drawback to positive trap technology is its defeatability by customers who obtain their own filters through theft, illegal purchase or construction. Another drawback is the loss of resolution in the secured channel's video caused by the filter's effect in the center of the video passband. Preemphasis is added at the headend to correct for the filter's response, but loss of picture content in the 2–3 MHz region of the baseband video signal remains. New positive trap schemes take advantage of sharper *surface acoustic wave* (SAW) or *crystal filter* technology. This approach allows the interfering carriers to be positioned a few kilohertz away from the secured channel's video carrier in contrast to 2.25–2.75 MHz with the conventional approach. Illegal construction of these filters is much more difficult. The new technology has the promise of achieving greater signal security and a higher quality picture for the authorized customer, but at a higher price.

Negative trapping removes signals from the cable drop to the customer's home. The trap is needed for customers who do not subscribe and is the least expensive means of securing a channel when over half the customers subscribe. The negative trap is ideal. There is no picture degradation of the secured channel because the trap is not in the line for customers who take the service. A drawback occurs for customers who do not subscribe to the secured service but want to view adjacent channels. These customers may find a slightly degraded picture on the adjacent channels due to the filter trapping out more than just the secured channel. This problem becomes more significant at higher frequencies, due to the higher Q (efficiency) required of the filter circuitry. From a security standpoint, it is necessary for the customer to remove the negative trap from the line to receive an unauthorized service. Maintaining signal security in negative trapped systems depends upon ensuring that the traps remain in the drop lines.

Scrambling and Addressability

There are two classes of scrambling technologies: RF synchronization suppression systems and baseband scrambling systems.

The concept of addressability should be considered separately from the scrambling method. Non-addressable converter/descramblers are programmed via internal jumpers or a semiconductor memory chip called a programmable read only memory (PROM) to decode the authorized channels. These boxes' authorizations must be physically changed by the cable operator. Addressable converters are controlled by a computer-generated data signal originating at the headend either in the vertical blanking interval (VBI) or by an RF carrier. This signal remotely configures the viewing capabilities of the converter. IPPV technology is supported by addressable converter/descrambler systems.

RF Synchronization Suppression Systems

Converter-based scrambling systems that perform encoding and decoding of a secured channel in an RF format comprise the commonly used scrambling technology. There are two basic RF scrambling formats. The more common is known as *gated* or *pulsed* synchronization suppression. With this method, the horizontal synchronizing pulses (and with some manufacturers, the vertical synchronization pulses) are suppressed by 6 dB and/or 10 MHz dB. This is done in the channel's video modulator at the intermediate frequency (IF). The descrambling process in the converter/descrambler occurs at its channel output frequency. This is accomplished by restoring the RF carrier level to its original point during the horizontal synchronization period. Variations of this method pseudo-randomly change the depth of suppression from 6 dB to 10 dB or only randomly perform suppression.

The other less popular format is known as *sine-wave* synchronization suppression. In this format, the scrambling effect is achieved by modulating the video carrier signal with a sine wave. This causes the synchronization to be suppressed as well as changes the characteristic content of the basic video information. This encoding process is performed at the IF frequency in the channel's video modulator at the headend. The decoding process in the converter/descrambler is accomplished by modulating the secured channel with a sine wave of the same amplitude and frequency as in the headend, but whose phase is reversed.

A phase-modulated RF scrambling technique based on precision matching of SAW filters constructed on the same substrate has been introduced. This low cost system is extending operators' interest in RF scrambling techniques for use within addressable plants.

Baseband Scrambling Systems

Baseband converter/descrambler technology provides a more secure scrambling technology for delivering video services. The encoding format is a combination of random or pseudo-random synchronization suppression and/or video inversion. Because the encoding and decoding are performed at baseband, these converter/descramblers are more complex and more expensive.

Maintenance of the system's video quality is an ongoing issue. The encoders are modified video processing amplifiers; they provide controls to uniquely adjust different facets of the video signal. The potential for set-up error in the encoder, in addition to the tight tolerances that must be maintained in the decoders, has presented challenges to the cable operator.

A recent extension to baseband video scrambling involves the addition of digitized and encrypted audio to the video signal's vertical and horizontal synchronization pulse intervals. While there are issues of Broadcast Television Systems Committee (BTSC) stereo compatibility, providers of these products claim that the combination of video scrambling with audio encryption provides the highest degree of security afforded by the industry.

Off-Premises Systems

The off-premises approach is compatible with recent industry trends to become more consumer electronics friendly and to remove security-sensitive electronics from the customer's home. This method controls the signals at the pole rather than at a decoder in the home. This increases consumer electronics compatibility since authorized signals are present in a descrambled format on the customer's drop. Customers with cable-compatible equipment can connect directly to the cable drop without the need for converter/descramblers. This allows the use of all VCR and TV features.

Signal security and control in the off-premises devices take different forms. Nearly all off-premises devices are addressable. Specific or all channels are controlled remotely. While there were several attempts to take modified addressable converter/descramblers and enclose them on the pole, this approach was not successful. The system was costly and less consumer friendly than having the converter/descrambler in the home because it delivered only a single channel at a time to the entire residence. With the growth of multiple TV and multiple VCR homes, this is an unacceptable limitation.

Interdiction technology involves a scheme similar to that of positive trap technology. In this format, the pay television channels to be secured are transported through the cable plant in the clear (not scrambled). The security is generated on the pole at the subscriber module by adding interference carrier(s) to the unauthorized channels. An electronic switch is incorporated allowing signals to be turned off. In addition to being consumer electronics friendly, this method of security does not degrade the picture quality on an authorized channel.

The Signal and the Customer's Equipment

The common devices at the cable drop to the customer's home are grounding safety devices called *ground blocks* and a two-way signal splitter that sometimes has a built-in grounding terminal.

Some systems use ground blocks or two-way splitters that incorporate a high-pass filter. These filters are used in two-way plant to minimize RF ingress into the cable plant in the 5–30 MHz reverse (upstream) spectrum. These filters have a low enough crossover frequency not to cause group delay in the downstream video channels.

Splitters or ground blocks should have little effect on picture quality provided there is adequate signal to handle the splitter's loss. The signal strength may be below specifications due to an excessively long drop or more activated cable outlets in the house than the cable design anticipated. To compensate, some systems use an alternating current (ac) powered drop amplifier. These amplifiers can create problems—a reduced CNR or increased distortions.

Consumer electronics switching devices, designed to allow convenient control and routing of signals between customer products and cable systems' converters/descramblers, have built-in amplification stages to overcome the losses associated with the internal splitters. These amplifiers add distortions or noise. When most cable systems were designed, consumer electronics switching devices were not taken into account because they did not exist.

Signal splitting in VCRs can be a problem. To compensate for recording SNR deficiencies, inexpensive VCRs sometimes split the signal unequally between the by pass route and the VCR tuner. This gives the VCR a stronger signal than the TV receiver to improve VCR performance. In addition, this strategy reduces the quality of the signal routed to the TV. When it is compared with VCR play-back, the disparity in performance is reduced.

Consumer Electronics Compatibility

Formed under the sponsorship of EIA and the NCTA, the JEC brings together technical representatives from the cable and consumer electronics industries to find ways to make cable service and consumer electronics products more compatible. It was soon recognized that the cable subscriber and the consumer electronics customer were one and the same. In order to better serve that customer, cooperation between the two industries was necessary.

For decades the two industries had evolved slowly but without coordination. As technology began to accelerate, a serious divergence became apparent. It was the consumer who ultimately suffered. The high-minded goal of improving the situation for the consumer ran into a few practical problems.

On the consumer side, costs were very critical issues. Additional expense is hard to recover in higher prices because of the extreme competitiveness of the electronics industry. That competitiveness is a consequence of the ability to produce larger quantities of very reliable, high quality products than the marketplace can easily absorb. This over capacity makes price almost impossible to sustain, much less raise. The nature of technology is such that it is impossible to contain it. Any technical advance quickly spreads throughout the industry, making differences between most brands difficult for the consumer to distinguish.

On the cable side, technology made possible experimentation in new services and the expansion of channel capacity. Many of these experimental services failed. Because these services were offered in set-top boxes owned by the cable operator, the consumer was protected from financial loss when service experiments end up in disappointment. Additionally, the basis of subscription service was the ability to deny that service to those who choose not to pay. If service is easily available to all, few would pay. Without payment, programming cannot be purchased and those who create that programming cannot be compensated. So signal security is critical to the delivery of subscription services.

Early on, it was clear that three things were needed to resolve much of these difficulties:

- A channel plan vs. frequency
- Tuner specifications
- A decoder interface.

A major feature of cable television is its self-contained spectrum. Because this spectrum is isolated from the spectrum in the environment, the usual restrictions placed on broadcast television do not apply. The entire spectrum can be filled with television signals. Early cable practice utilized a variety of methods for labeling these channels. This diversity made it difficult to use a common labeling scheme for consumer electronics products with extended tuning range. A frequency plan was necessary to allow common nomenclature.

These cable signals can be very well controlled in their relative strength. This is in sharp contrast to broadcast signals. To prevent broadcast co-channel interference, channels are spaced. The VHF and UHF bands have gaps where others have claim on the spectrum. It is common for the broadcast tuner to attempt to tune a distant weak signal centered between nearly adjacent strong local signals. Thus the demands on tuners designed for broadcast and for cable are very different. It is difficult to make a tuner that performs well in both situations.

A most serious difference in the requirements of tuners designed for cable and those designed for broadcast is the issue of direct pick up. A broadcast tuner needs minimal shielding; the spectrum being tuned is only occupied by the broadcast signal. A cable tuner has another problem. Because the propagation speed through cable is about two thirds that of free space, the broadcast signal arrives at the TV first. A few microseconds later, the cable signal arrives. If the tuner directly responds to the over-the-air signal, the two signals will be mixed and interference will result. The solution to this problem is covered by a now-expired U.S. patent, the Mandell patent (3,333,198, July 25, 1967). The cable signal is selected by a tuner specifically designed for the purpose. It is well shielded and has the ability to discriminate against immediately adjacent channels. One constraint is relaxed. Signal levels across the tuned band differ only slightly in level. This tuner's output is converted to a single channel chosen to be one not used locally. Therefore, there is no off-air signal to be directly picked up by the TV and mixed with that delivered by the cable system. The problem of direct pick up is solved only because the output of the converter is a single channel on a frequency not used locally.

The use of the set-top converter solves one problem but makes the TV's tuner useless. Those features which depend on the tuner are also made inoperable. In particular, the remote control that came with the TV or VCR is only used to turn those devices on and off. The set-top's remote control changes channels. In the case of the VCR, the timer that changes channels is crippled because it cannot change the channels of the converter. Taping different channels consecutively becomes impossible.

In order for a cable ready TV or VCR to provide acceptable performance when connected to cable, it has to have a minimum amount of shielding to make it immune to direct pick up interference. Also, it must not radiate spurious signals up the cable. These signals can come from the local oscillator or from the harmonics of all of the digital signals which provide on-screen displays and drive other features. These signals can cause interference to other receivers in the same residence or in adjacent residences if they are allowed to enter the cable. All of these issues have been covered by a cable RF specification.

Cable's requirement for signal security caused the need for set-top boxes to move beyond mere frequency conversion. It took on the function of descrambling of signals in the early 1980's. If a cable ready TV or VCR were designed with adequate tuning range, adequate shielding against direct pick-up and the ability to handle a full spectrum of adjacent signals without overload, it might still require a set-top descrambler. This would duplicate the circuitry in the TV, provide slight additional signal degradation and interfere with certain features built into TVs and VCRs. The solution to this problem was the decoder interface connector. This was a 20-pin connector on the back of TVs and VCRs that brought video and audio signals out of the consumer electronics product for external processing. The descrambled signals were returned to the consumer electronics product for display or recording. The Decoder Interface connector was given the name *MultiPort* by the EIA because it was envisioned to have multiple uses. The main advantage of the MultiPort was that it made scrambling nearly transparent to the user. Cable subscribers loved it. Somewhere between 300,000 and a million TVs were produced with the plug. Several thousand descrambler plug-ins were made. However, only certain segments of the cable industry promoted the MultiPort. Much of the rest of the industry experienced so few complaints from subscribers that they didn't feel motivated. Additionally, many of these cable operators were enjoying a healthy revenue stream from the rental of remote controls. Since the MultiPort didn't need an additional remote control, that revenue stream was threatened. Another practical problem was locating consumers who bought MultiPort equipped sets. The consumer electronics industry lost patience. They did not promote the feature and most consumers who acquired these products had no idea what they had purchased. To this day, there are several hundred MultiPort plug-ins still in cable operator inventories. Knowledgeable installers call this to the attention of subscribers when they stumble across an old MultiPort TV. This results in a happy customer.

In an attempt to improve upon and formalize the MultiPort, the JEC created a standard that was adopted by the American National Standards Institute (ANSI): the ANSI/EIA 563 Decoder Interface Standard. Even though it was a closely related cousin to the MultiPort, this standard was never implemented in products and was retired.

In 1992, Congress passed the Cable Act which has far reaching implications for cable system operation and economics. An amendment to the Cable Act requires the FCC to create rules to improve the compatibility between consumer electronics products and cable operations. These rules regulate the conditions under which scrambling can be employed and provide a technical definition for cable-ready products. That definition included the channel plan created by the JEC, specifications on tuner performance and the inclusion of a decoder interface connector. Products that do not comply with the Commission's technical specifications cannot be sold as cable ready, cable compatible or under any other term which might mislead consumers.

Under these rules, cable operators will be required to provide plug-in modules for those subscribers who have cable-ready products and request the modules. Cable service will have to be compatible with the decoder interface.

The technical details of the decoder interface are many and complex. They are found in the interim standard EIA-105.1 and EIA-105.2. The first describes the physical interface and the second the minimum protocols needed to operate the system. The decoder interface supports digital plug-ins that will allow analog receivers to be used for the reception of digital signals at standard resolution.

As complex and difficult as the decoder interface has been, the DTV future presents even more interesting challenges for the consumer interface. The still embryonic technology is evolving at a rapid pace taking advantage of the fact that we can now afford to put well over a million transistors into consumer products to process signals. Additional millions of transistors can be included in memory elements. This tremendous processing power offers a wide variety of options for digital video compression and new interactive and MultiMedia services. The biggest unknowns center on the consumer.

The very nature of digital systems accommodates the ability to do something in software and to allow digital components to be reconfigured electronically. Thus, closely related but not necessarily identical approaches can be handled with little or no extra cost. This was not easily possible in the analog world. As more functionality is absorbed into software based systems, the need for a single strictly enforced standard is diminished. Multiple approaches and self-adaptation will ease the burden.

The 1996 Telecommunications Act included a section requiring the FCC to create rules pertaining to the retail availability of set-top terminals. The FCC issued a *Report and Order* in June 1998 (FCC 98-116, CS Docket No. 97-80). The decoder interface is an appropriate choice for these needs with analog signals.

Another issue becomes important as we move to subscriber ownership of the electronics that implement services (in most cases, the set-top box). In the old paradigm, the service provider answered questions and supplied training in the use of the equipment. The equipment was repaired at no additional cost. If the service failed to enjoy sustaining support, the cable operator absorbed the equipment loses. In the new world of subscriber ownership, the retailer and manufacturer of the equipment must stand ready to answer questions on usage and provide training and repair. The consumers' expectations must be adjusted to accept situations in which the equipment becomes no longer useful because the service failed or was replaced with a improved version. The cable operator will no longer be absorbing the losses since the equipment is no longer his.

An additional complication arises. If subscriber-owned set-top boxes introduce interfering signals into the cable system, they will have to be disconnected. If the subscriber insists on using them, that subscriber will have to be disconnected. There will be charges for service calls to solve these problems. Consumer will be angry when the hardware they purchase cannot be used on the cable system because of interference it may cause.

The Hybrid/Fiber Coax Architecture

A cable system can be considered to be made up of three main parts: trunk, feeder and drop. The trunk is intended to cover large distances, up to ten miles. When the trunk is made of coaxial cable, amplifiers are required about every 2,000 ft (.6 km). Surprisingly, the trunk portion of the cable plant only covers about 12% of the total footage. The feeder portion of the cable system supports taps for subscribers. It has a maximum length of about a mile and a half. This limitation comes from the fact that energy is tapped off to feed homes. Consequently, the power levels must be relatively high. These higher power levels reach into the slightly nonlinear regions of the amplifiers. As a result, only a few amplifiers can be used before the distortion begins to negatively impact the picture quality. Approximately 38% of the total footage in a cable system is in the feeder portion of the plant. The drop is the flexible cable which goes to the home. It has a maximum length of 400 ft (122 m). Typically, it is more like 150–200 (45.7–61 m)ft in length. Approximately half of the total footage of cable in a system is in the drop and the flexible wiring in the home.

The feeder portion of the cable plant is a hotbed of activity. Every day new subscribers are added and others are removed. Approximately 20% of Americans move to a new residence every year. When they move out, the drop is usually disconnected. When they move to their new residence, the drop must be re-connected. It is important that the technology used in the feeder portion of the trunk supports this constant activity. It must be "craft-friendly"; that is, it must be easy to work in. The trunk portion of the cable plant is relatively stable. There are very few changes made.

The lasers that drive the fibers are expensive, costing many thousands of dollars. The receivers that convert the optical energy into signals TVs and VCRs can use are also relatively expensive. To be practical, these components must serve hundreds of subscribers each so that the costs can be shared. This is accomplished by fibers which feed small coaxial systems, which in turn serve a few hundred to a couple thousand subscribers.

The HFC architecture has made it possible to cost effectively increase bandwidth, signal quality and reliability, while reducing maintenance costs and retaining a craft-friendly plant. It makes two-way service practical. The bandwidth of coaxial cable has no sharp cut off. It is the cascade of amplifiers which limits the bandwidth. Twenty to forty amplifiers in cascade not only reduce bandwidth but also constitute a severe reliability hazard. Overlaying low-loss fiber over the trunk portion of the plant eliminates the trunk amplifiers. This in turn leaves only the distribution portion of the plant with its relatively short distances and only two or three amplifiers. Wider bandwidth is thus facilitated. Two-way operation becomes practical for two reasons. First, the fiber itself is not subject to ingress of interfering signals. Secondly, the cable system is broken up into a large number of small cable systems, each isolated from the others by its own fiber link to the headend. If ingress should cause interference in one of these small cable systems, that interference will not impair the performance of the other portions of the cable plant. The Time Warner Quantum cable system has been in operation for several years at 1.047 GHz and 150 analog channels.

Today it is common practice to install passive components (taps, splitters, amplifier housings) which will pass 1 GHz. Actual 1 GHz amplifiers are still too costly. However, some systems are being built "GHz ready" so that the addition of GHz amplifier modules will upgrade the system with little waste of installed plant. While GHz capability is a future possibility, 750 MHz upgrades are commonplace today.

The fiber backbone approach breaks the cable system into a multitude of much smaller cable systems with amplifier cascades limited to four to six amplifiers. Each of these small cable systems is fed with a fiber link to the headend (see Figure 6.13-7). The advantages include significantly lowered vulnerability to amplifier outages, reduced bandwidth restrictions and noise build-up due to amplifiers in series and greatly reduced ingress. The latter effect makes two-way cable practical. A major attraction of the fiber backbone is that its implementation cost is low. Fiber is over lashed onto the existing trunk plant. The in-place cable is broken into segments and used for the small-scale cable systems. Some of the amplifiers are reversed in direction. Nothing is wasted.

If we now turn our attention to the distribution plant, we find that we must run higher signal levels to support the tapping of energy to serve drops to customers. The higher signal levels mean that we begin to reach into nonlinear areas of the amplifier's operating characteristic. Nonlinear distortion builds up. In addition, the taps are not perfectly impedance-matched to the cable. Consequently, the signal is reflected back and forth between the taps, resulting in a smearing of the picture.

Rogers Cable of Canada has suggested the answer to these difficulties. They have called their technique "Super Distribution." In one form of its implementa-

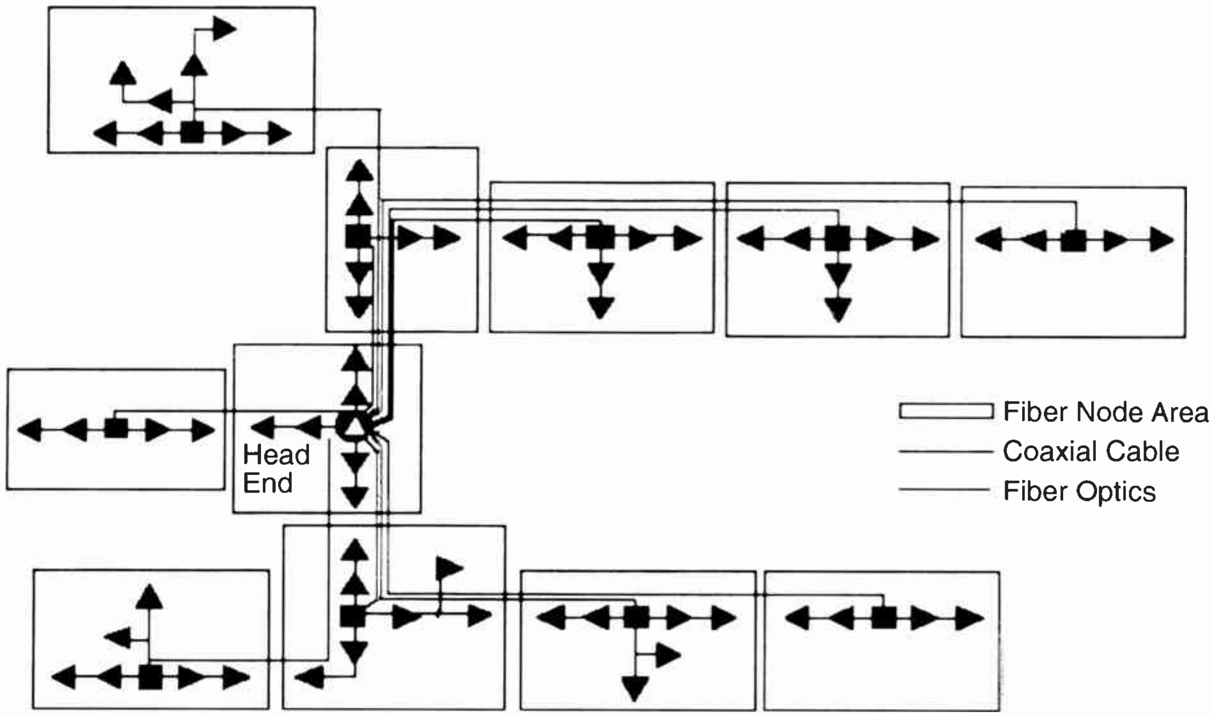


Figure 6.13-7. Cable systems with fiber optic backbone trunk.

tion, line-extender amplifiers are structured to have up to three hybrid amplifier chips. One feeds the next line extender amplifier, one feeds half of the taps back to the previous line extender and the third feeds half of the taps to the next line extender. The existing tapped feeder cable is cut in half between line extenders. New, untapped cable is overlashed to connect line extenders. The consequences are that the signal level between amplifiers is lower since that cable is not tapped. The signal level on the tapped runs is lower because they are shorter. Nonlinearities are reduced, and they do not build up as in the previous structure. Also, signal leakage may be less of a problem because of the lower signal levels. In addition, the number of taps in series in any cable is drastically reduced, thereby reducing the amount of micro-reflections experienced by any one subscriber. This technique effectively cures the ills of the distribution portion of the plant (see Figure 6.13-8).

With fewer amplifiers in series, the constraints on their design and operation are reduced. Higher bandwidths become practical.

Compression

The interest in digital television goes back several decades. Long before digital television was practical for consumer use, it held interest for military and professional applications. The principal attractions are those of all digital signals: the ability to transmit signals over arbitrarily long distances and to store the signals without degradation, the ability to process signals for security and other purposes, the ability to remove unnecessary redundancy to increase transmission and storage efficiency, and the ability to add appropriate coding and redundancy so signals can survive harsh transmission environments. Even as early as the late 1960s, work on digital television for broadcast and consumer applications was underway. At that early date, Zenith Radio Corporation worked on the digitization of television signals for the purpose of enhancing scrambling for an over-the-air subscription television service.

Compression along with digital modulation has given massive capacity to the cable spectrum. Compression gained its first industry wide momentum in the pursuit of HDTV—also called Digital Television (DTV). The analog bandwidth of signals from an HDTV camera consists of approximately like 30 MHz each of red, blue and green. Nearly 100 MHz of analog signals would require more than a gigabit per second for straight digital transmission. The FCC, which sets federal regulations on spectrum usage, only allows 6 MHz! As a result, there is a need to remove a lot of the redundancy in the picture. The same technology can be used to squeeze multiple standard definition signals into the same 6 MHz. When Compression is used to put multiple current quality signals into 6 MHz, the result is called standard definition television (SDTV).

It is commonly felt that movies may eventually be digitized effectively at a 1.5 Mbps rate while live video may require 3 Mbps to 5 Mbps. Data rates currently used in trials are somewhat higher because of the preliminary nature of the equipment. When more ad-

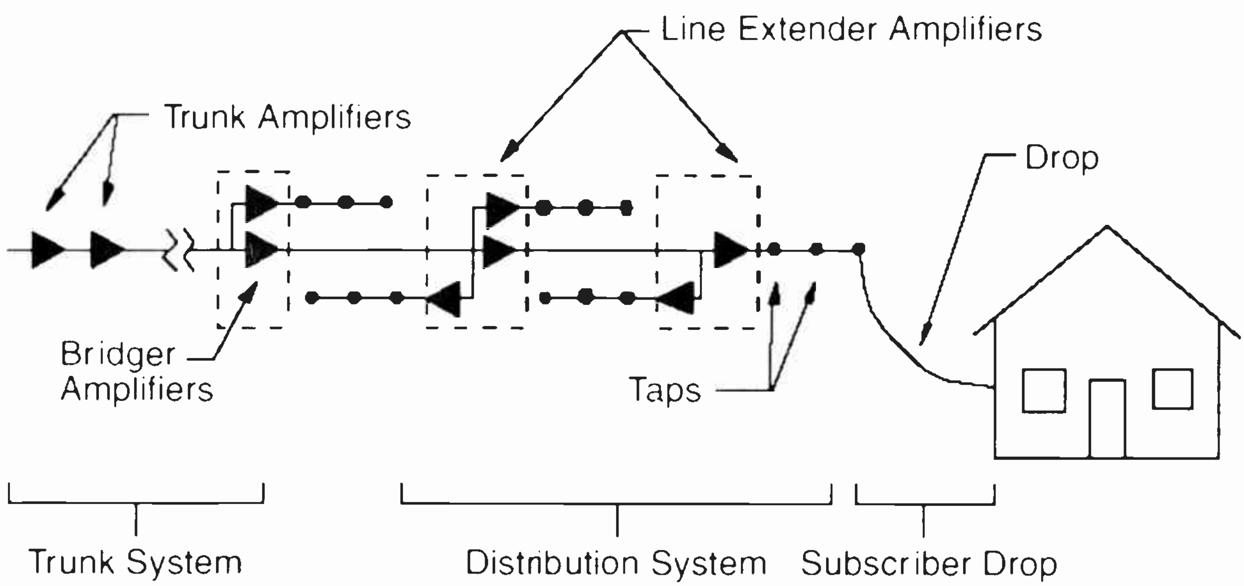


Figure 6.13-8. Super distribution plant.

vanced digital video compressors become available, data rates should be lowered at least a little more. The reasons for the lower data rate with movies stems from two main factors. First, movies have a 24 frames per second rate while video has 30 frames per second. Motion reproduction in video is better than in movies, but there is a 20% increase in data required for the same image quality. Secondly, movies can be processed in non-real time. This means that they can be iteratively processed. The compressor parameters can be adjusted on a scene by scene basis. Much more processing can be brought to bear since the computations do not have to be done in real time.

Digital Transmission

There are two distinct aspects to the digitization of video and audio. The first is called *Source Coding*. Source coding is the conversion of the source information into digital form and its processing to meet certain objectives. The objectives include the amount of resolution, the tolerance for artifacts, the maximum data rate, and the cost constraints. All of these objectives dictate the kind of processing that makes sense for any particular application.

Channel Coding tailors the signal for the medium used. The characteristics of satellite, broadcast, cable, microwave, twisted pair, fiber optics, video disc, magnetic tape, etc. are quite different. Some media are essentially noise free. Other media have little random noise, but extensive bursty noise. Still other media are plagued with multipath and co-channel interference while media exist with essentially none of these difficulties. These varieties of characteristics yield different requirements for modulation and error protection, which are the fundamental techniques of channel coding.

A signal may experience multiple channel codings and re-codings. For example, the signal may start out on video tape at a satellite uplink using the type of channel coding appropriate for magnetic media. It might then be conveyed via satellite to a broadcast affiliate using channel coding appropriate for a satellite link. It could then be broadcast over the air using channel coding for that environment. Finally, it might be picked up by a cable system headend and put through the cable system using channel coding designed for fiber optics and cable. Each of these links would use a different modulation method and error protection scheme designed to optimize performance in its environment. The source coding remains the same; the channel coding changes to fit the media. Since the spectrum in the cable plant is very well behaved, two or more broadcast source coded signals might be combined in one 6 MHz cable channel.

Modulation serves the same purpose for digital signals as it does for analog signals. While digital signals can conveniently utilize time division multiplexing, frequency division multiplexing remains an important part of the plan. There are several reasons for this. Practically, digital circuits are limited in their speed. They cannot support a TDM structure for the entire GHz spectrum of cable. Breaking the spectrum into 6 MHz pieces relaxes the speed requirement on the digital circuits. TDM can be utilized within each of the 6 MHz channels.

Additionally, FDM facilitates the transition from analog television to digital television. No cable system is known to have plans to convert all signals to the digital format. Those cable systems which are adding digital, are doing it for those customers who want more. They are adding it to the current collection of analog channels. In many cases, the analog scrambling is left in place. The investment in analog set top boxes is too great to abandon. By separating the spectrum into two areas, one for analog channels and one for digital channels (including TDM), a rational transition to digital technology is accomplished.

Unfortunately, the modulation methods chosen for use on cable (QAM) and for over-the-air transmission are different. The ATSC standard selected a Vestigial Sideband Modulation (VSB) method with a reference carrier. This was done because the broadcast environment must contend with co-channel interference between multiple digital television signals, between digital television and NTSC, and between NTSC and digital television signals. The VSB method has been specifically designed for this environment. The reference carrier aids the television tuner in acquiring the signal when the channel is changed.

Two VSB structures were created. The over-the-air 8-VSB structure has a payload of almost 20 Mb/s and can carry one HDTV signal with relative robustness. Since cable has a more disciplined spectrum with fewer challenges, a 16-VSB version was created which has less error immunity and double the data capacity. It can carry two HDTV signals.

The Advanced Television Test Center (ATTC) completed tests of 16-VSB on cable in Charlotte, North Carolina. The 16-VSB system tested delivered 38.5 Mbps in all locations. Some older cable systems were included in the tests. The longest amplifier cascade was forty-eight! Fiber links and microwave Amplitude Modulated Links (AMLs) were also included. 256-QAM is expected to be able to deliver similar results. Thus it is possible to transmit twelve compressed movies in 6 MHz at 3 Mbps.

Current thinking is to allocate the lower end of the down stream cable spectrum to analog signals. This covers the 50 MHz to 450 MHz (or 550 MHz) frequencies. A basic unscrambled tier of broadcast signals would be at the low end of the spectrum. At higher channel numbers, trapped pay services or analog scrambled services would be supplied. Above 450 or 550 MHz, digital signals would be provided to the band edge located at 750 MHz or 1 GHz. Thus, from 200 MHz (33 six MHz channels) to 550 MHz (91 six MHz channels) will be available. With twelve movies per 6 MHz, 396 to 1092 movies may be offered simultaneously. Of course, digitized video signals (rather than movies) may also be offered by reducing the total number of programs.

OpenCable™

CableLabs, the cable industry's research and development consortium has created a number of purchasing specifications on behalf of the industry. These are not standards, since CableLabs is not an accredited standards setting organization. Rather, these specifications define the functionality and external parameters of products used in cable but leave the internal details to the manufacturers. For example, CableLabs will not specify a single microprocessor or operating system for these applications. But, to insure interoperability, CableLabs establishes a certification program to verify that the products comply with the specification and will perform satisfactorily.

OpenCable™ is one such specification. The goal for OpenCable™ is to create a multi-vendor interoperable environment of digital set top terminals. OpenCable™ is intended to facilitate a range of new interactive services and function that are more like a digital set top computer than just a digital set top terminal. An important objective is competitive supply of all of the elements of the system including the advanced set top terminal, its operating system, core applications and plug-ins, conditional access system, platform management system, billing, provisioning and subscriber management systems.

The procedure began with a Request For Information, RFI, sent to leading computer and consumer electronics companies. Twenty-three companies responded. From these submissions, CableLabs solicits vendors to author specific elements of the suite of specifications. An Intellectual Property, IP, pool is created with cross licensing of all IP required for the specification. Implementation IP not inherently required by the specification is left to vendors to license in their normal fashion.

Key OpenCable™ interfaces include the formats for digital cable television, consumer privacy systems, copyright protection system, interfaces for high-speed connection to the Internet and interfaces required to author interactive applications.

A *Harmony Agreement* for Digital Standards has been achieved. Five open layer standards, including MPEG-2 video, the Dolby Digital audio AC-3 system, MPEG-2 transport multiplex, ATSC service information tables, and the International Telecommunications Union ITU-J.83 Annex B modulation system are included in the agreement. The modulation system calls for 64 quadrature amplitude modulation (QAM) carrying 27 Mb/s and 256 QAM carrying 40 Mb/s with concatenated trellis code plus enhancements such as variable interleaving depth for low latency in delay sensitive applications. All of these agreements have been endorsed by CableLabs and most have been adopted as digital standards by the Society of Cable Telecommunications Engineers.

OpenCable™ will use an implementation of the federally published core data encryption standard (DES). By utilizing common core encryption, operators will not have to duplicate the transmission of programming and related content in order to implement dual conditional access. While operators would still need to maintain separate headend control systems to control each supplier's set top terminals, the conditional access data can be carried in a single unified bit stream.

CableLabs will implement a vendors certification process where by specific products can be tested for compliance with the OpenCable™ specifications. The certification process will allow these devices to be made available to consumers through retail outlets.

CableLabs recommends that its member utilize the specifications for their procurements. However, members will make their own procurement decisions. This is not a joint purchasing agreement.

The Digital Video Standard for Cable

The Society of Cable Telecommunications Engineers, SCTE, is an accredited American National Standards Institute, ANSI, organization which utilizes the due-process methods required by ANSI to set standards for the cable industry. This is the mechanism by which Digital Video for use on cable is standardized. These SCTE standards are an integral part of OpenCable™.

This Digital Video Transmission Standard for Cable Television describes the framing structure, channel coding, and channel modulation for a cable television realization of a digital multi-service distribution system. The standard applies to signals which can be distributed via satellite directly to a cable systems, as is the usual cable practice. The specification covers both 64 and 256 QAM.

The modulation, interleaving and coding is tailored to the characteristics of cable system practice in North America. The modulation is selected by the cable system to be either 64-QAM or 256-QAM. The forward error correction approach utilizes a concatenated coding approach that produces low error rates at moderate complexity and overhead. The system goal is one error event per 15 minutes. The input data format is MPEG-2 transport.

Two modes are supported: Mode 1 for use with 64-QAM with a symbol rate of 5.057 Ms/s and Mode 2 for use with 256-QAM with a symbol rate of 5.361 Ms/s. Higher order QAM may be considered in the future as warranted by experience.

The cable channel, including optical fiber, is a bandwidth-limited linear channel, with a balanced combination of white noise, interference, and multi-path distortion. The QAM technique used, together with adaptive equalization and concatenated coding is well suited to this application and channel.

The transport layer for MPEG-2 data is comprised of fixed length packets having 188 bytes. One byte is for synchronization purposes followed by three bytes of header contain service identification and scrambling and control information. This is followed by 184 bytes of MPEG-2 or auxiliary data. The data stream is transmitted serially with the MSB first. The first packet byte is transmitted as a checksum which is replaced by a sync byte having a fixed value of 47HEX by the QAM demodulator.

Interleaving is used to enable the correction of burst noise errors.

Competition

The combination of HFC and compression have dramatically and completely changed the course of cable's evolution. An important reality which must be appreciated in creating the next vision of cable's future is that technology cannot be contained. The nature of technology is that it is available to all who wish to apply it. Thus the HFC architecture has been adopted by many in the telephone industry as a practical approach. "Fiber to the home" has essentially been abandoned. Likewise, compression is available to all who wish to compete with cable. From the consumer's perspective, it's the programming that matters. The delivery mechanism and the industry delivering the programming are secondary as long as certain minimum standards of signal quality, reliability, and cost are maintained.

One of cable's main advantages in the past has been massive capacity due to a wider bandwidth than is available to most potential competitors. Compression makes much of that advantage available to many potential competitors by more efficient utilization of whatever bandwidth those competitors have available.

The cable franchise is almost always non-exclusive. Those who consider cable to be overly profitable were always able to gather up investors and build a second or third cable system and compete. This has been attempted in a number of situations only to result in financial losses in most cases. It's only human nature to believe that "The grass is always greener on the other side." There have only been a few instances of success following this course of action.

Except for building a second cable system, there really was no effective method of deploying direct competition to cable before compression. Cable's high bandwidth gave it so much programming that other delivery means were anemic in comparison. Compression changes that. Now it is possible for a competitive service to be launched with sufficient programming to be a reasonable alternative even with less available actual bandwidth.

Video delivery competition comes from several sources:

- Direct Broadcast Satellite
- Telephone plant delivering video
- Alternative cable plant builders
- Microwave Multipoint Delivery System (MMDS)
- Broadcast
- Physical distribution

The vision of cable in the late 1990s has to include the reality that government regulation will play a bigger role. That role will limit what can be done and how it can be financed. It will also impact the way competitors behave and change the balance between cable and its opponents.

The Video Services

There appear to be three kinds of visions for video service under discussion:

1) Video Electronic Publisher
2) Video Common Carrier
3) Video Internet

These three visions vary considerably in how practical they are and how closely they approach reality.

The Video Electronic Publisher is the model of traditional cable television. The cable operator selects programming thought to be of interest to subscribers and packages it in convenient form. A selection of local off-air channels, satellite delivered programming, and local origination is divided into tiers. If the cable operator selects and packages programming which interests the subscriber, the reward is acceptance and profits. If the programming does not please the subscriber, the result will be financial difficulties.

This service is a broadcast service in that all subscribers get essentially the same programming. A possible exception is that in some cases, special ethnic programming is delivered on a neighborhood by neighborhood basis. This takes advantage of the fact that several different trunk lines leave the cable headend and go to diverse geographic areas. The signaling is mostly one-way. The only need for two-way is for the ordering of Impulse Pay Per View (IPPV) programs. There are over a hundred and sixty programs on satellites which could be packaged and delivered to subscribers. Other ideas for new channels arise continuously. In the analog world, there is only capacity to carry about eighty total channels. After local broadcasters and other mandated channels are accommodated, there is room for only around sixty of these satellite delivered programs.

This cable service consists of several levels. Basic Service includes all of the mandated broadcast channels plus the Public, Educational, and Government (PEG) channels required by the local franchise. The 1992 Cable Act requires that these channels be unscrambled. A group of tiers can then be created out of the satellite delivered channels. The tiers can be as simple as just one cluster or a set of clusters based on genre of programming. When a set of channels is offered on an individual channel basis, that service is called "a la carte". At the next level are pay channels such as HBO or Showtime. These are premium channels with a substantially higher price. They are sold on a monthly subscription basis. Next are IPPV channels which sell individual program viewings. These consist mostly of movies, but also include concerts and sporting events. A special case of IPPV is True Video On Demand (TVOD) or Near Video On Demand (NVOD) to be discussed below in more detail. Then there are the new MultiMedia and interactive services which are considered part of the National Information Infrastructure (NII) or the Information Superhighway.

The 1992 Cable Act has given broadcasters new rights. Broadcasters are allowed to choose between *must carry* status or *retransmission consent* status. Must Carry status means that the broadcaster's signal must be carried on the channel number the broadcaster selects. That selection is made from those channel numbers occupied on two specified dates or the broadcaster's own channel number. Selection of Must Carry status foregoes any right to compensation for the signal. Alternatively, broadcasters can choose Retransmission Consent status which precludes cable carriage without the consent of the broadcaster. The broadcaster and the cable operator negotiate on the terms. Broadcasters can change their status every three years.

The Video Common Carrier

The Video Common Carrier model is sometimes called video dial tone. As can be assumed from its name, this is the kind of service proposed by many telephone operators. While originally fiber to the home was proposed, all such projects now involve an HFC architecture. Like the Video Electronic Publisher model, this model has a portion of its programming which is essentially a broadcast service. Everyone gets the same signals in this part of the Video Common Carrier service. The main difference is that the video dial tone provider does not have any editorial control over what is presented. In principle, enough capacity is provided so that any programmer who wishes carriage can have it on a non-discriminatory basis. The Video Common Carrier model is mostly one-way down stream signaling with a small up-stream capacity for ordering programming.

The second portion of the Video Common Carrier model includes subscriber selection of programming either through a NVOD or a TVOD approach. TVOD implements the full promise of Video Dial Tone.

There are some proposals to use Asymmetric Digital Subscriber Loop (ADSL) technology with switching. A relatively low bit rate signal is switched onto the subscriber twisted pair copper loop based on a very low bit rate request signal sent to the point of origination. In this model, no common services exist. All services are specifically requested one at a time. Decompression converts the digital signal into a form suitable for display on consumers' TVs and VCRs.

The Video Internet

The Video Internet is the most visionary and least practical approach of all. It envisions a symmetrical capacity in both directions. Subscribers can obtain all the normal cable services plus they can originate programming and send it to those who would wish to view it. This is the most expensive of the three service models and has the least in common with the cable plant already installed.

The Video Internet requires digital video switching and symmetrical or near-symmetrical signal capacity. The costs of implementing this kind of service will be very high. Unlike the current Internet service, whose costs are partially government supported, the Video Internet would be expensive to use. It remains to be seen if consumers will find these services of enough value to be willing to pay the price suppliers would need to ask.

An interesting irony exists when digital video and digital voice services are compared. People watch video for hours and video requires mega (millions of) bits per second. People talk on the telephone for minutes and voice requires only kilo (thousands of) bits per second. The ratio of the number of bits needed for telephony compared to the number of bits needed for video is in the range of hundreds of thousands! If bits are bits and their transport and switching comes at a given cost, either digital video will be incredibly expensive and digital voice affordable or digital video will be affordable and digital voice essentially free! If this is not the case, then the "bits are bits" paradigm must be somehow violated.

TVOD versus NVOD

A major choice to be made is that between TVOD and NVOD. Here is where the practicality and the reality of the vision of the Information Super Highway come into stark contrast. The greatest unknown is what do the consumers want and what are they willing to pay.

TVOD is any video, any time, with full virtual VCR functions. That is the ability to do stop, fast forward, rewind, pause, slow motion, etc. It doesn't take too much to realize that this is essentially impossible. There must be some compromises. The question is how much compromise will be acceptable to the consumer? What is the consumer's desire in the trade off between cost and convenience?

NVOD is an engineering approximation to TVOD. It is based on the Pareto Principle which states that most of the desire for video will be for the most popular titles. Perhaps 80% of all demand will be for the top 20% of titles—perhaps for even fewer. So the first engineering approximation is to limit the number of titles. The next is to limit the frequency of access. Rather than allow everyone instant (or near instant) access to the video of their choice, the most popular movies will be stagger started on several channels. In the Time Warner Quantum system in Queens, New York, the top five movies were stagger started about every half hour. Thus, the maximum wait time for a top movie to start is a half hour. The average wait time is fifteen minutes. Since they were approximately two hours long, each of these movies required four channels. The Quantum cable system has 1 GHz of bandwidth and 150 analog channels.

With compressed digital video, it is possible to offer many more movies with much more convenient start times. The target for wait times in NVOD is to make most of them slightly less than the time to drive to the video store. An example can be illustrative.

Consider a cable system with a 750 MHz maximum frequency. The range of frequencies from 50 MHz to 450 MHz is allocated to ordinary analog television. (Below 50 MHz, the system is used for up-stream communications.) This 400 MHz of spectrum contains 66 analog 6 MHz channels. This would make a fine ordinary cable service.

The 300 MHz from 450 MHz to 750 MHz contains 50 channels of 6 MHz each. If we are able to compress movies into 3 Mbps data streams and if we use a modulation scheme such as 16 VSB which has a payload of 38.5 Mbps, we'll be able to carry a dozen movies in each 6 MHz frequency slot. Since there are 50 of these slots, there are a total of 600 synthetic channels for movies.

Consider offering the top ten movies (each approximately 135 min long) on twelve synthetic channels each. This makes the stagger time 11.24 min and occupies 120 total synthetic channels. Now put the next twenty most popular movies on eight synthetic channels each for a repeat time of 16.9 minutes. 160 channels will be used up. Put the next forty most popular movies on roughly half hour repeats (like in the Quantum system). That will take four channels each and occupy a total of 160 channels. Take sixty more movies and repeat them continuously in sixty more channels. We have 130 movie titles occupying 500 synthetic channels. This leaves us with 100 channels.

In the remaining 100 channels, all analog scrambled signals will be duplicated in the digital format so that a digital subscriber doesn't need any analog processing. This will likely take another twenty channels. This leaves eighty channels for continuous repeat of short subjects such as shopping channels dedicated to single categories each.

This provides a tremendously convenient and comprehensive service with no switching and no complex head end server. The head end is a rather straight forward play back center.

PC versus TV and Interactivity

There is a great deal of debate about what the consumer of the future will expect in a television set. Those from the computer industry want to build TVs into PCs and those from the consumer electronics industry want to build PCs into TVs. Of course, the whole world seems to want to put PCs into set-top boxes.

A few moments of thought about how these products are used should shed some light on what is the evolution likely to be supported by consumers. All one needs to do is recall how it feels to be in the family room watching TV with someone who has a remote control in hand and is flipping channels. Certainly, this is the lowest level of interactive television. Anything more interactive may not be appreciated by others in the room. The conclusion seems to be that interactive television is an individual activity.

Now consider the information packing density of the television screen. The NTSC television technical standard can just barely support text consisting of forty characters across and twenty rows. That is woefully inadequate for anyone who has used modern personal computers. The input device for a television set, the remote control, is also very limited. It is easy to conclude that the proper instrument for serious information highway work (or play) is the personal computer.

The user of the personal computer is accustomed to upgrading performance, to spending a few hundred dollars on software or hardware modules, and to monthly subscriptions to on-line services costing ten to twenty dollars. The average personal computer user spends thousands of dollars on the hardware and software and sees great utility from this. There may even be a motivation of providing learning opportunities for the children.

Contrast this with the average TV and VCR user. They have spent not thousands but mere hundreds of dollars for their equipment. That equipment has relatively low resolution, low speed, and has minimal other capabilities. These are the consumers who complained bitterly about monthly cable bills of around twenty dollars for basic and ten for premium service. The average TV lasts twelve to fifteen years and the average VCR about five. They are not upgradeable in any way.

The TV and VCR are likely to be used when the day is done and there is no more energy left and it's "wind down" time. For most people interested in or capable of "interactivity" the personal computer is where it's done. Even when these highly interactive people turn on their TV or VCR, it's because they have no more energy to interact with anything! They just want good passive entertainment.

The Electronic Program Guide, Interactivity for the Common Man

While we have concluded that the interactivity vision is primarily a personal computer activity, there are a limited number of very appropriate interactive applications the consumer will likely want for the TV receiver and VCR. By far the most promising is the electronic program guide (EPG). The EPG is most appropriate because it fits the TV and VCR application. It helps the consumer navigate through the multitude of channels. There is a consumer motivation to learn to use it.

An interesting form of EPG is the *mood guide*. It is based on the theory that there are a limited number of moods viewers have when they turn on the TV. Those moods can be studied and understood and programming categorized so that it is associated with the set of moods. When the TV is turned on, the first thing seen is a list of possible moods. The viewer reflects briefly and then indicates the mood of the moment. The EPG responds with all the programming which is appropriate to the moods of the viewer. It also suggests what may be available later for viewing or taping. Thus the likelihood that the viewing experience will be satisfying is increased.

Another important role for the EPG is to keep a record of which types of programs each family member watches. Over a period of a few months, enough information is gathered to allow the EPG to advise the viewer that programming consistent with past viewing patterns is available for viewing or taping. It becomes less likely that programming of interest will be missed. This increases the value of the subscription to the subscriber and increases retention. In general, the satisfaction of the consumer increases.

The EPG contains all of the information necessary for automatic operation of a VCR. The time, the channel, and the program category are available in digital form. The consumer's main responsibility is to ensure blank tape is loaded in the machine!

The EPG is interactivity for the common man.

The Fax Machine, E-Mail & the NII for the Common Man

In the early 1990s, e-Mail and the Internet have become important parts of the professional work place. Those who are enthusiastic about these modern communication tools envision them becoming part of ordinary daily life for nearly everyone in the late 1990's. This may be the case for a minority; but the reality is that the average consumer does not have the interest or the skills to work these systems.

However, there is a very cost effective device which is much easier to use and nicely demonstrates the Pareto Principle: The fax machine. The fax machine is rapidly becoming a consumer electronics product. There are a large number of models under $250 and the price is rapidly approaching $100.

Even more importantly, the fax machine is easy to use. It does not require any complicated ability to navigate through menus or complex computer commands. The fax machine requires minimal skills to operate. It requires no touch typing. Nearly everyone can operate the fax machine. The Information Super Highway for the fax machine already exists: the phone system.

In many cases, the fax machine can send a page at less than the cost of postage. Certainly, the speed of delivery is greater and more assured. When e-Mail and other alternatives are considered, the fax machine is an extremely cost effective and easy to use alternative.

The fax machine is e-Mail and the National Information Infrastructure for the common man!

Section 7: Signal Measurement and Testing

Introduction by Randy Hoffner, ABC, New York, NY

The rapid advances of digital technology for radio and television applications have redefined the requirements of signal measurement and analysis. Contrary to popular belief, digital is not perfect. Problems still develop. Such problems, in fact, are often more difficult to identify than their analog counterparts. Within this environment, of course, is the continued requirement for maintenance of existing analog devices and systems. These demands are being met with advanced test instruments and built-in diagnostic systems for new hardware. Despite the powerful new tools available to broadcast engineers today, the greatest need continues to be the training of maintenance personnel. Machines do a very good job of taking measurements, and a rather poor job of interpreting them. In this section, both instruments and procedures are examined.

"The ability to quantify audio signals in terms of characteristics and qualities is paramount in audio engineering." Stanley Salek simplifies audio measurement in his chapter entitled, *Analog Audio Signal Analysis.*

Handbook Editor-in-Chief, Jerry Whitaker, prepares engineers to move "full speed ahead into the digital domain" in his *Video Signal Analysis* chapter.

The section concludes with Donald Markley's analysis of radio frequency signals.

7.1 ANALOG AUDIO SIGNAL ANALYSIS

STANLEY SALEK, P.E.
HAMMETT & EDISON, INC., CONSULTING ENGINEERS, SAN FRANCISCO, CA

AUDIO MEASUREMENT

The ability to quantify audio signals in terms of characteristics and qualities is paramount in audio engineering. Comparisons can be made with reference to established standards and requirements, including measurements relating to amplitude, frequency content, distortion, noise and phase. The observation of such attributes allows a virtually complete characterization of an analog electrical audio system.

In the broadcast environment, audio measurements are used to gauge the overall quality of equipment such as amplifiers, recording systems, mixing consoles, digital audio devices and other networks throughout the overall broadcast signal path.

Amplitude Measurement

The most basic of needs in audio measurement is to determine a value relating to the size, or amplitude, of an audio signal. Since an audio waveform is rapidly changing, methods have been developed to convert peak, root mean square (RMS) and average values of the changing waveform into corresponding proportional dc voltages that can be more easily observed.

There are specific cases where the peak value is the most direct measure of magnitude. It gives an indication of the largest excursions (either positive or negative) of an audio waveform. As shown in Figure 7.1-1, the audio signal is applied to an absolute value circuit, which rectifies the waveform such that the output is all positive. A diode is then used to couple into C and R. These serve as memory and decay time elements, respectively, that can be adjusted in value to conform to the ballistics desired. Although the output is still changing with time, along with the input, the excursions corresponding to the peak values of the original waveform are much slower and more easily observed on metering devices. As the value of resistor R is increased, the decay time of the output is proportionally increased as well. If the resistor is completely removed, a peak hold circuit results.

Peak (actually peak-to-peak) functions can also be observed on an oscilloscope, although this technique is often impractical because of the difficulty in reading the random waveforms typical in most audio material. Storage oscilloscopes can perform a peak-to-peak hold function.

While there are many cases where the peak value is of considerable use, the RMS value of a signal is generally most meaningful since it gives indication of the energy content of the signal without regard to its waveform. In audio measurement, however, it is usually simpler to detect the average amplitude of a given waveform and relate it to an associated RMS value, with reference to a sine wave. The RMS level can be defined as follows:

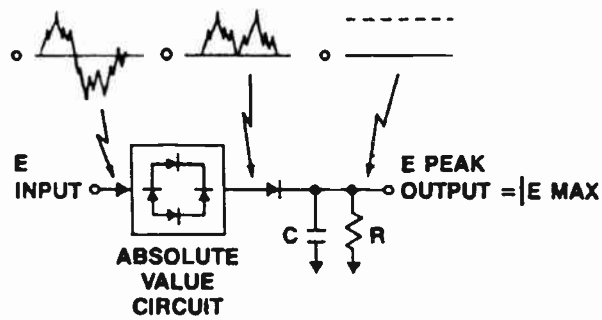


Figure 7.1-1. Peak value detection.

$$E_{rms} = \sqrt{\frac{E_1(t)^2 + E_2(t)^2 + \cdots + E_n(t)^2}{n}}$$

where E_1 through E_2 are successive measurements over a total of n samples. As can be seen from its name, the value is computed by taking the average of n samples of E squared. Performing the square root function completes the calculation. This function is also commonly referred to as *true RMS*.[1] Figure 7.1-2 shows how this is accomplished electrically.

Through the use of the absolute value circuit of Figure 7.1-1 and the R/C configuration found in the RMS detector of Figure 7.1-2, an average detector can

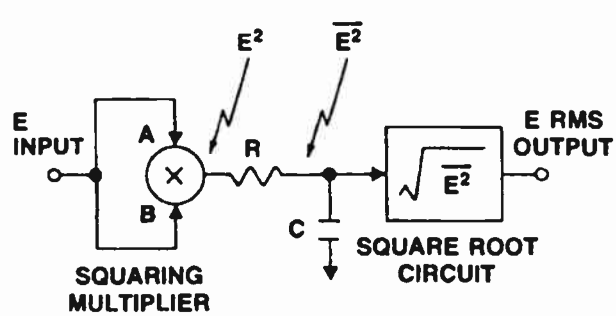


Figure 7.1-2. True RMS detection.

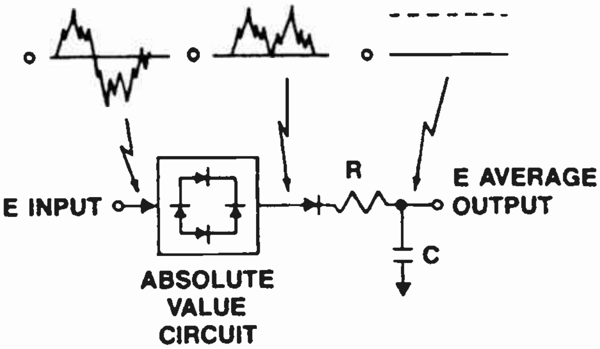


Figure 7.1-3. Average value detection.

be made, as shown electrically in Figure 7.1-3 and mathematically:

$$E_{average} = \frac{E_1(t) + E_2(t) + \cdots + E_n(t)}{n}$$

In terms of audio perception, the average value of an audio signal is related to program material density, where the peak value described earlier relates to a maximum. Since the peak value defines the upper limit of allowable modulation in a transmission system, it is often technically desirable that the peak-to-average ratio be as low as possible to attain highest perceived loudness and signal-to-noise ratio (SNR). Achieving this effect may require compromising aesthetic goals and may not always be appropriate, depending on the type of program material.

The decibel (dB), is a unit for comparing relative levels of voltage or power signals in transmission systems. In broadcast audio systems, the most common representation of decibels is dBm. This is the value of a signal with reference to 1 mW into a 600 Ω load. The level in dBm of a signal can be found using the following relation:

$$\text{dBm} = 20 \log\left(\frac{E}{0.775}\right)$$

where E is the voltage level to convert. The number 0.775 represents the voltage level reference of 0 dBm. Note that, strictly speaking, this formula is only true when the circuit impedance is 600 Ω. In practice, the formula is used typically without regard to the impedance level. Voltage levels obtained from the peak, RMS, and average circuits described previously can be used for possible values of E. When this is done, some common types of metering can be synthesized to observe the activity of audio material.

Metering

Two popular types of metering for the characterization of program audio are the *standard volume unit indicator*, commonly known as the *volume unit* (VU) meter, and the *peak program meter* (PPM). Although VU metering has been more common in U.S. broadcast equipment, the standard PPM indicator is often found as well, especially in audio mixing consoles.

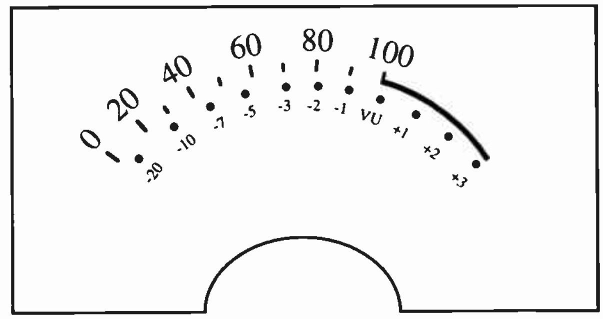


Figure 7.1-4(a). VU meter.

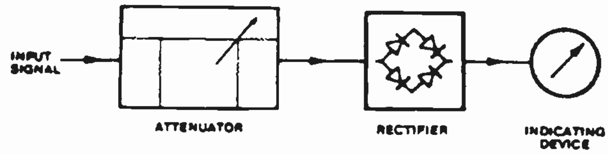


Figure 7.1-4(b). Block diagram of the stages of a typical volume unit indicator.

The VU meter was introduced in 1939 to serve as a standard program level indicating device (see Figures 7.1-4(a), 7.1-4(b)).[2] Its original purpose was to be the reference between broadcasters (as well as other programming suppliers) and the telephone company. A VU meter is the combination of a bridge rectifier, a resistive attenuator and an ammeter with an approximately voltage linear scale to produce an average responding ac voltmeter. The VU meter is calibrated such that it reads 0 VU across a circuit in which a sinusoid develops 1 mW in a resistance equal to the circuit impedance (0.775 volts RMS across 600 Ω). This allows the meter to be powered directly by a 600 Ω program line, with the attenuator typically set to read 0 VU at +8 dBm.

Beyond reading continuous tones, VU meter dynamic characteristics are set so that it will display 99% of its steady state reading on a sine wave tone burst 300 ms long, with a fall to 5% of the reading in 300 ms. Essentially an average responding device, the VU meter will not respond to short duration program peaks. Therefore, levels normally should be set with a 10 dB margin (headroom) before the point of clipping.[3]

The PPM is designed to read nearly the full peak value of the audio signal (see Figures 7.1-4(c), 7.1-4(d)). It uses a rectifier and an integrator, producing a fast rise and slow fall effect on the display device. Typical standards require the PPM to read −2 dB, ±0.5 dB of the steady state value for a tone burst of 10 ms, and take 2.8 seconds for the pointer to fall 20 dB.[4]

Typically, a PPM exhibit flats frequency response and is calibrated such that the nominal peak program level corresponds to a 0 dB meter reading near full scale (generally +16 dBm).

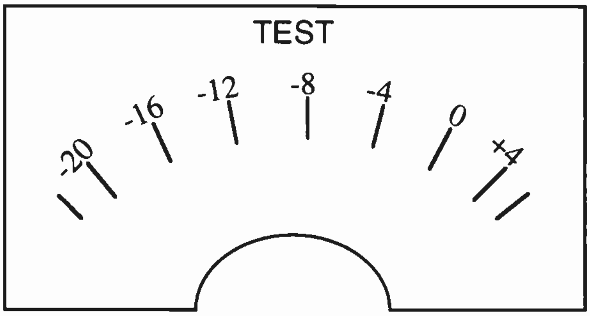


Figure 7.1-4(c). Arrangement of a typical peak program meter scale.

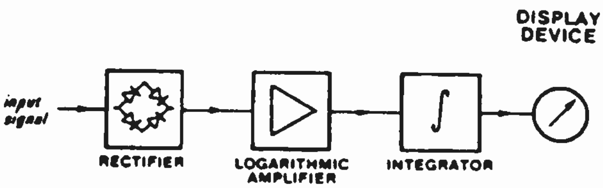


Figure 7.1-4(d). Block diagram of the stages of a typical PPM.

Other types of metering devices have been developed to read wideband audio in a simultaneous mode. One such system consists of an LED bar graph display exhibiting peak program content. Riding upon this is a brighter display (utilizing the same display elements) corresponding to VU standards. Such an indicator allows continuous monitoring of program material compression and dynamic characteristics. Another system combines a VU movement meter with peak indicating LED flashers. A *hold* function is sometimes associated with these flashers to allow an operator sufficient time to observe the approximate peak content.

FREQUENCY ANALYSIS

Amplitude analysis methods, as described in the previous section, are generally used to provide an indication of signal levels simultaneously over the entire audio range. It is sometimes more desirable, however, to be able to measure discrete frequencies in an audio system, allowing frequency response measurement as well as dynamic measurement of energy content throughout the audio spectrum.

Simply stated, frequency response is the capability of a device or system to pass or amplify equally, all frequencies within a specified range.[5] As far as audio in the broadcast environment is concerned, the range of interest is generally 50 Hz to 15 kHz or 20 kHz. Although few musical instruments produce fundamental frequencies greater than 4 kHz and the human voice much above 1 kHz, the reproducing device or system must be able to pass the harmonics that accompany the fundamental frequencies. Without adequate bandwidth or with uneven frequency response, an unnatural *coloration* of the perceived sound becomes evident. To overcome this potential problem, a great deal of care is taken to construct amplifiers with very flat frequency response to high frequencies. Since the responses of series connected amplifiers are additive, care must be taken to verify the flatness of each device in a system.

Several methods are available to measure audio frequency response. They include *discrete frequency measurement* and *swept frequency methods. Parallel analysis* and *Fast Fourier transform* (FFT) techniques also can be used.

The discrete frequency measurement method is uncomplicated and inexpensive. A simple measurement system consists of a low distortion audio frequency oscillator and a wideband ac voltmeter. A typical audio frequency oscillator is shown in Figure 7.1-5. The oscillator output is connected to the input of the device or system to be characterized. The voltmeter is used to observe the level at the output of the device, or at a desired intermediate point in a system.

The measurement is done by first setting the generator output level to the nominal input operating level of the device. Generally, a 400 Hz or 1000 Hz frequency is chosen initially in high fidelity audio systems. The device output level is read on the ac voltmeter, and this quantity is noted as a 0 dB relative reference. Provided the generator itself has a flat frequency response, measurements at frequencies through the audio band can be taken while recording the corresponding dB output levels with respect to the reference. A convenient and commonly used technique is to increment the frequency in a 1, 2, 5 sequence (20 Hz, 50 Hz, 100 Hz, 200 Hz, etc.). This method permits plotting the final response data on 4-cycle LOG/LIN graph paper, providing regularly spaced frequency increments horizontally. The logarithmic amplitude data are plotted along the horizontal axis, with the zero dB relative reference placed in a convenient position on the linear vertical axis.

Although the discrete frequency measurement technique is straightforward, it is also often tedious and time-consuming. Numerous frequency measurements must be made to ensure adequate testing. This method is most usable in the response measurement of single ended devices that do not have a suitable input port for connection to an audio generator. Transcription equipment, such as compact disc players are examples of these types of devices. Test recordings supplied by the equipment manufacturer and other sources are used to provide the tones necessary for discrete frequency response characterization.[6,7]

Swept Frequency

A faster and more efficient means of measuring frequency response is the swept frequency method. This process employs a sweep frequency generator as a signal source and measures response over the entire range of interest in one sweep. The detector for these measurements is most often a tracking type that follows the signal source and measures a narrow band of frequencies centered around the source frequency. Use

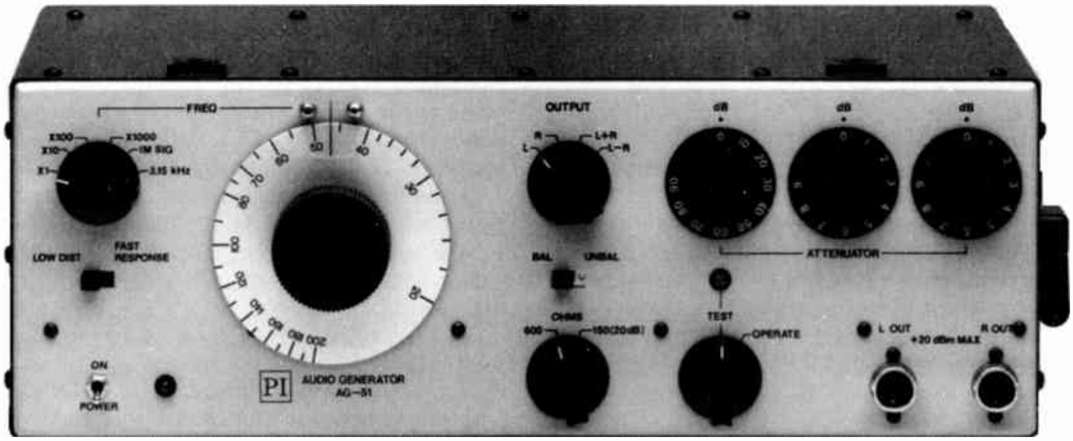


Figure 7.1-5. Audio frequency generator designed for broadcast use.

of a tracking detector is a better guarantee that the amplitude measured is that of the tone generator and is not influenced by spurious noise or harmonics.

Devices specifically designed to conduct swept frequency measurements include wave analyzers and spectrum analyzers. Wave analyzers, which are no longer commonly available, must be used with plotters to provide a hard copy data plot, while spectrum analyzers directly produce response images on a built-in display. A representative spectrum analyzer, which is specifically intended for audio frequency use, is shown in Figure 7.1-6.

Figure 7.1-7 shows a typical setup for measuring amplifier frequency response using the swept frequency method. The signal source used to drive the test device is the tracking oscillator output of the analyzer. The device output is terminated with an appropriate characteristic load impedance and connected to the analyzer input. Measurement of the frequency response is made by manually or automatically sweeping the analyzer across the frequency range of interest. A plotter, a scope-type camera, or a printer (if the analyzer supports one) can be used to provide a permanent record of the test device response characteristics.

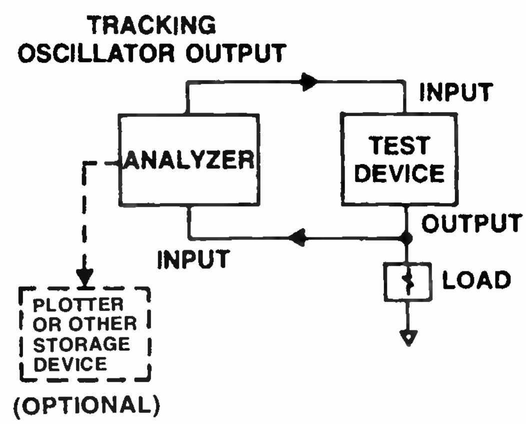


Figure 7.1-7. Frequency response measurement using swept method.

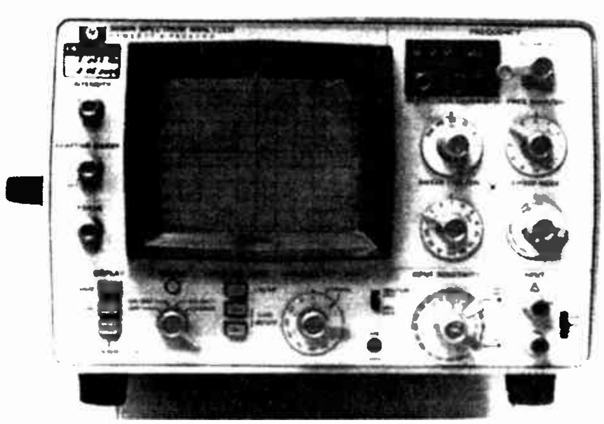


Figure 7.1-6. Spectrum analyzer intended for audio use.

The swept frequency method also can be used to measure the overall frequency response of multiple-track audio recorders. The tracking oscillator signal is connected to the record amplifier input and the output is taken from the playback amplifier output, allowing recording and playback frequency response measurement to be done all at once. However, a time delay between input and output signals may exist, caused by the physical displacement between record and playback heads in analog recurring systems. This effect can be corrected simply by using a very slow sweep speed or a wider measurement bandwidth.

Real Time Analyzer

A *real time audio analyzer* (RTA) consists of a sequential collection of one octave or one-third octave filters having individual detectors and indicators at each output. The program audio is simultaneously fed to the inputs of all the filters. The output signal of each filter is proportional to the amount of energy

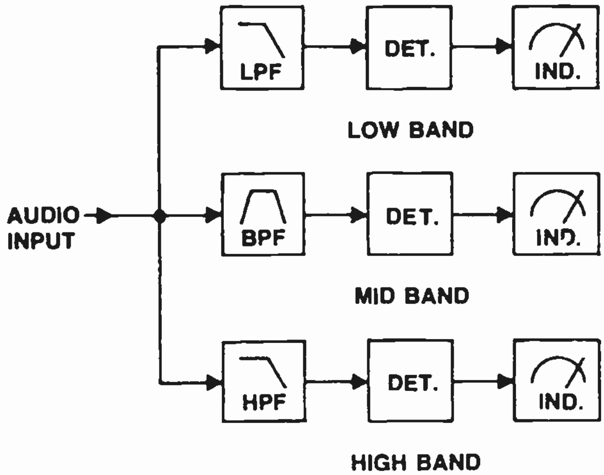


Figure 7.1-8. Basic real-time analyzer.

occurring in that particular frequency band. This technique is also referred to as *parallel analysis.*

A simplified version of an RTA is presented in Figure 7.1-8. As shown, it is intended to break the audio band into three sections using low-pass, band-pass and high-pass filtering. Signal detectors are then used to condition the audio for display on a suitable indicator, one set for each of the three bands. The detectors can be (and often are) the same peak, RMS or average circuits described earlier. Typical readout indicators are bar graph displays with dB-calibrated scales. When arranged side by side, the readouts provide a graphical presentation of amplitude versus frequency, similar to spectrum analyzer display. Unlike the spectrum analyzer, however, an RTA does not rely on a fixed sweep speed.

Parallel techniques using the RTA are often used for dynamic program material and room acoustics analysis.[8] This type of analyzer is also useful for measuring frequency response of audio devices when used in conjunction with a *pink noise* source. Pink noise has a constant mean squared voltage per octave of frequency. This characteristic makes it popular in audio work, since it allows correlation between successive octaves by ensuring the same voltage amplitude is available as a reference. By connecting the pink noise source to the input of a device to be characterized and the RTA to its output, a response curve can be displayed almost instantly. The characteristic response of pink noise is shown in Figure 7.1-9, as compared with *white noise.* White noise is unweighted, since its response is flat with frequency, except that it is attenuated for frequencies above the audio spectrum.

Although often prohibitive in cost, network and FFT analyzers are also exceptionally useful in audio frequency domain measurements. Network analyzers are swept analysis instruments, used to characterize two-port networks (devices having an input and output) as to frequency, phase and delay responses. They are employed where substantial accuracy in the measurement of these parameters is required. RF subsystem

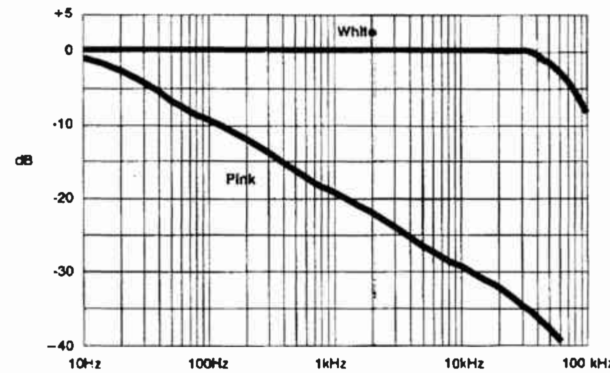


Figure 7.1-9. Response versus frequency for white noise and pink noise.

and semiconductor device design have been the major application for network analyzers, although some newer generation equipment includes audio-frequency coverage. In the case of audio systems, network analyzers allow precision response measurement of amplifier and filter designs.

FFT analyzers have the ability to convert a snapshot sampling of an audio or other time-varying source and mathematically transform the result into a display of the frequency components present. Because the conversion process is done by a specialized digital signal processing (DSP) microcomputer, an FFT analyzer often can produce a complete spectrum display as much as an order of magnitude faster than conventional swept spectrum analyzers. FFT analyzers are most helpful in low frequency measurement, where a swept analyzer would require a very slow sweep time to resolve closely spaced components. As their costs have been decreasing, FFT analyzers are becoming increasingly popular for audio system analysis and measurement.

DISTORTION MEASUREMENT

When a two-port device is driven beyond its range of linear operation or through areas of discontinuity, signal distortion occurs. As a result, additional frequencies appear at the device output that were not present at its input. In cases where distortion becomes extreme, it can be identified through listening. Odd-order distortion (such as clipping distortion) can become audible at around 1.25%. Even-order distortion, characterized by a coloration of the program material, becomes audible at about 5%. Generally, systems with a wider frequency response capability need to maintain lower distortion levels to be acceptable. Since distortion is not always obvious to many people, techniques are available to measure its various types.

Classic audio distortion can be characterized in two basic ways: *harmonic distortion* and *intermodulation distortion.* While the two associated methods produce uncorrelated measurement values, each gives a quantitative result of device quality in terms of a single number. Although total harmonic distortion (THD) content is determined by only one method, intermodulation distortion (IMD) has several accepted measurement practices, including the SMPTE and CCIF methods. *Transient intermodulation* (TIM) distortion is another commonly measured type of IMD. Digital audio signals can be characterized using THD and IMD methods, but they also are susceptible to other types of distortion.

Harmonic Distortion

Harmonic distortion is a measure of individual harmonic amplitudes with respect to the amplitude of the fundamental frequency. In practice, harmonics greater than third order often add little to the resultant value because of their negligible amplitude. THD is defined as:

$$\text{THD\%} = 100\frac{\sqrt{A_2^2 + A_3^2 + A_4^2 + \cdots + A_n^2}}{A_1} \quad (4)$$

where A_2 through A_n are the amplitudes of the individual harmonics and A_1 is the amplitude of the fundamental.

As seen in Figure 7.1-10, a 1 kHz sine wave with harmonic distortion shows only minor differences when it is overlaid with an undistorted signal (as viewed on an oscilloscope). The amplitude and slope errors do not lead directly to a numeric result. But when a spectrum photo of the same waveform is observed (Figure 7.1-11), the above relation can be applied. With the fundamental at 0 dBm (0.775 V), the second harmonic at -26 dBm (38.8 mV), and the third harmonic at -50 dBm (2.5 mV), the harmonic distortion can be calculated:

$$\text{THD\%} = 100\frac{\sqrt{(0.0388)^2 + (0.0025)^2}}{0.775} = 5.0\% \quad (5)$$

Although spectrum analysis can produce accurate THD measurement results, a simpler and more cost effective procedure that produces a direct numeric quantity is more popular. Figure 7.1-12 shows the block diagram of a typical THD analyzer. An oscillator (with much less harmonic distortion than the device or system to be measured) is connected to the test device input. The distorted output signal of the device is filtered to remove A_1, the fundamental component. This produces a signal that, when RMS detected, is proportional to the THD imposed by the device being tested.

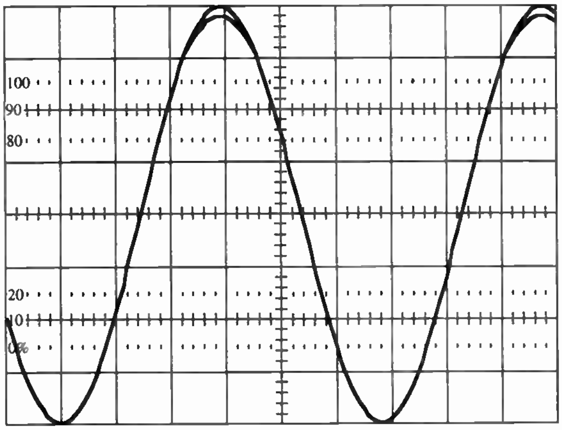


Figure 7.1-10. Comparing the distorted output of an amplifier with its undistorted 1 kHz input component.

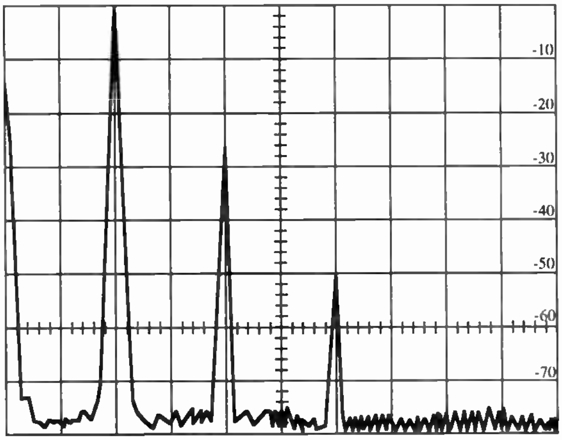


Figure 7.1-11. Measuring THD with a spectrum analyzer (V: 10 dB/div; H: 500 Hz/div).

THD measurement is often conducted using the same 1, 2, 5 sequence of frequencies mentioned for discrete response measurement. The THD results can be plotted on the same graph to characterize the device under test on a single page. THD measurements may be taken over various input levels, but as the level is reduced, noise characteristics may affect the readings. In such cases, the spectrum analyzer method could produce more meaningful results.

Intermodulation Distortion (IMD)

The intermodulation method of measuring distortion uses a test signal composed of two sinusoidal signals of different frequencies (except for TIM measurement, to be covered shortly). After summation, they produce the effect of an amplitude modulated carrier when applied to a circuit having IMD. The intermodulation method is useful because the harmonic distortion of the signal sources do not affect the measurement.

The SMPTE method uses a low frequency (f_1) and a relatively high frequency (f_2) signal (usually 60 Hz and 7 kHz, respectively) that are mixed at a four to one amplitude ratio (see Figures 7.1-13 and 7.1-14). This method involves the measurement of the relative amplitude of the modulation sidebands added to the higher frequency signal. For diagnostic purposes, it is often useful to determine even-order and odd-order distortions separately, although this is best done by spectrum measurement techniques. Even-order distortion usually can be characterized by the ratio of the sum of the amplitudes of only the two second-order

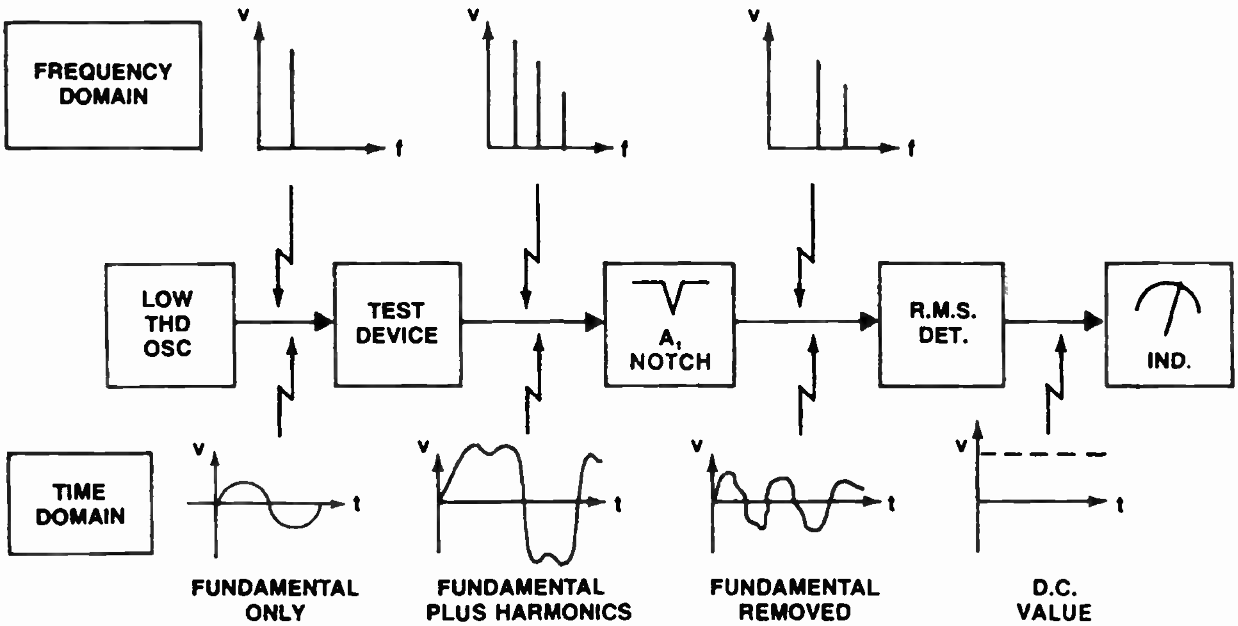


Figure 7.1-12. THD analyzer.

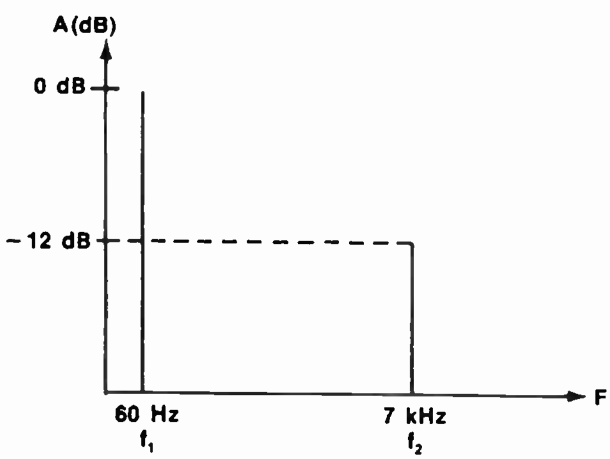


Figure 7.1-13. Spectrum of SMPTE IMD input test signal ratios.

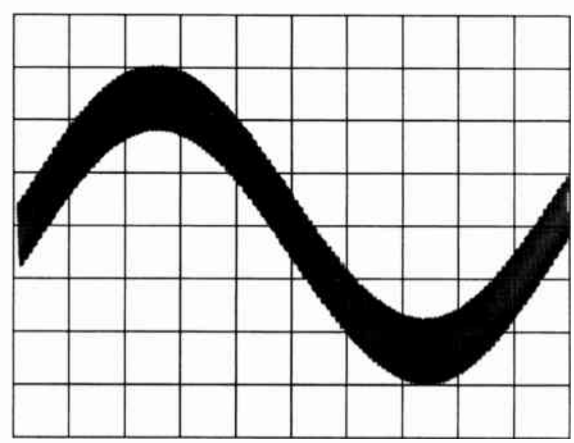


Figure 7.1-14. SMPTE IMD test signal as viewed on an oscilloscope.

spurious frequencies, $f_2 - f_1$ and $f_1 + f_2$, to the amplitude of the carrier signal, f_2

$$\text{SMPTE IMD\%} = \left[\frac{A_{(f_2-f_1)} + A_{(f_1+f_2)}}{A_{f_2}}\right] \times 100 \quad (6)$$

(second order)

In a similar manner, odd-order distortion can be characterized by the ratio of the sum of the amplitudes of the two third-order spurious frequencies, $f_2 - 2f_1$ and $2f_1 + f_2$ to the amplitude of f_2:

$$\text{SMPTE IMD\%} = \left[\frac{A_{(f_2-2f_1)} + A_{(2f_1+f_2)}}{A_{f_2}}\right] \times 100 \quad (7)$$

(third order)

Figure 7.1-15 shows the output signal of an amplifier with IMD, as viewed on an oscilloscope. Note the elongated trough as compared to Figure 7.1-14. As with THD, spectrum analysis can be used to determine the numerical amount of distortion present. Intermodulation sidebands can be seen around f_2 in the spectrum photo of Figure 7.1-16. Second- and third-order distortion percentages for this example are calculated as follows:

$$A_{f_2} = -12 \text{ dBm} = 195 \text{ mV}$$

$$A_{(f_2-f_1)} = A_{(f_1+f_2)} = -38 \text{ dBm} = 9.76 \text{ mV}$$

$$\text{SMPTE IMD\%} = \left[\frac{9.76 + 9.76}{195}\right] \times 100 = 10.0\%$$

(second order)

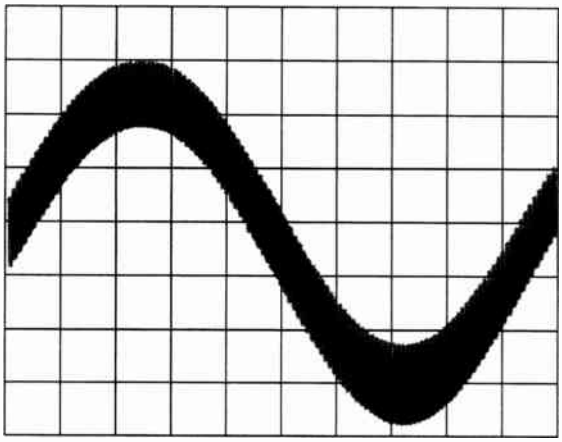


Figure 7.1-15. Output of an amplifier exhibiting significant SMPTE IMD.

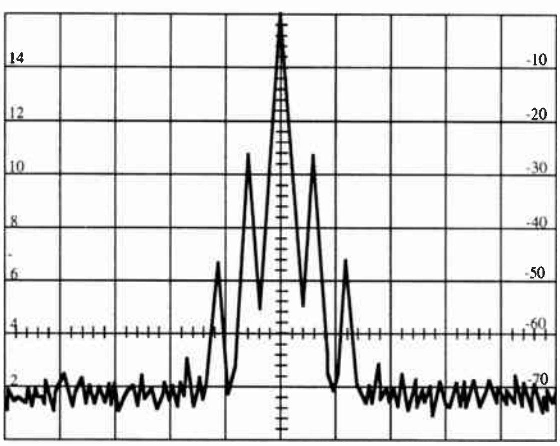


Figure 7.1-16. SMPTE IMD measurement. (V: 10 dB/div. Top of screen −12 dBm; H: 100 Hz/div. Center frequency: 7.0 kHz).

$$A_{(f_2-2f_1)} = A_{(2f_1+f_2)} = -58\ \text{dBm} = 0.98\ \text{mV}$$

$$\text{SMPTE IMD\%} = \left[\frac{0.98 + 0.98}{195}\right] \times 100 = 1.0\% \tag{8}$$

(third order)

As is shown, the contribution of even-order distortion products is usually greater than that of the odd-order. To express the result as a single quantity, the vector sum of the two quantities is taken:

$$\text{SMPTE IMD\%} = \sqrt{(\text{IMD\%Even})^2 + (\text{IMD\%Odd})^2} \tag{9}$$

(total)

$$= \sqrt{10^2 + 1^2} = 10.05\%$$

As with THD, SMPTE IMD has a direct method of numeric solution, as shown in the block diagram of Figure 7.1-17. The two test frequency oscillators are summed to produce the $f_1 + f_2$, signal, which is then applied to input of the device to be tested. The distorted output signal is high-pass filtered to remove the f_1 fundamental component, leaving only the amplitude modulated f_2 component. Using a standard AM demodulator and low-pass filter, the residual f_1 component is obtained. After RMS detection, a dc level proportional to the distortion is produced that can be viewed on a direct reading indicator.

Wow and flutter is a term that describes a special case of IMD normally associated with analog audio tape recorders. It is caused by variations in tape velocity across the recording and/or reproducing heads, due to imperfections in the mechanical drive system. These variations result in frequency modulation of the recorded and reproduced signal. The frequency spectrum obtained is similar to that of the SMPTE IMD measurement method, except the f_1 low frequency signal is generated by fluctuations in tape speed and is not of any set amplitude.[5]

To measure wow and flutter, a test tape containing a prerecorded 3 kHz (or 3.15 kHz) tone is played. Using an audio spectrum analyzer, for example, the amplitude of the first sideband (A_m) with reference to the 3 kHz amplitude (A_0) is measured. The frequency of the flutter (F_m) also must be known. Then the following relation can be used to approximate the percentage of wow and flutter present:

$$\text{WOW \& FLUTTER\%} = \frac{2(A_m)(F_m)}{(A_0)(3\ \text{kHz})} \times 100 \tag{10}$$

(peak-to-peak)

For example, if the first sideband (either upper or lower) amplitude is 10 dB below the 3 kHz maximum amplitude (0.316 V with reference to an arbitrary 1 V amplitude at 3 kHz) and is at a frequency of 4 Hz, the wow and flutter would be 0.084%.

The CCIF intermodulation method uses a combination of two higher frequency sinusoidal signals (f_3, f_4) of equal amplitude. They are typically 1 kHz apart and found at 5/6 kHz, 14/15 kHz, or 19/20 kHz in many applications. One of the spurious frequencies generated is low in frequency while others are gathered around the two driving frequencies. Figures 7.1-18 and 7.1-19 spectrally show the driving frequencies before and after passing through a test amplifier. As with SMPTE IMD measurement, the generated spurious products can be classified as even-order or odd-order. Even-order distortion is expressed as the ratio of the amplitude of the difference component ($f_4 - f_3$) to the sum of the two driving frequencies (f_3, f_4):

$$\text{CCIF IMD\%} = \left[\frac{A_{(f_4-f_3)}}{A_{f_3} + A_{f_4}}\right] \times 100 \tag{11}$$

(second order)

Odd-order distortion is determined by calculating the ratio of the sum of the amplitude of the two third-

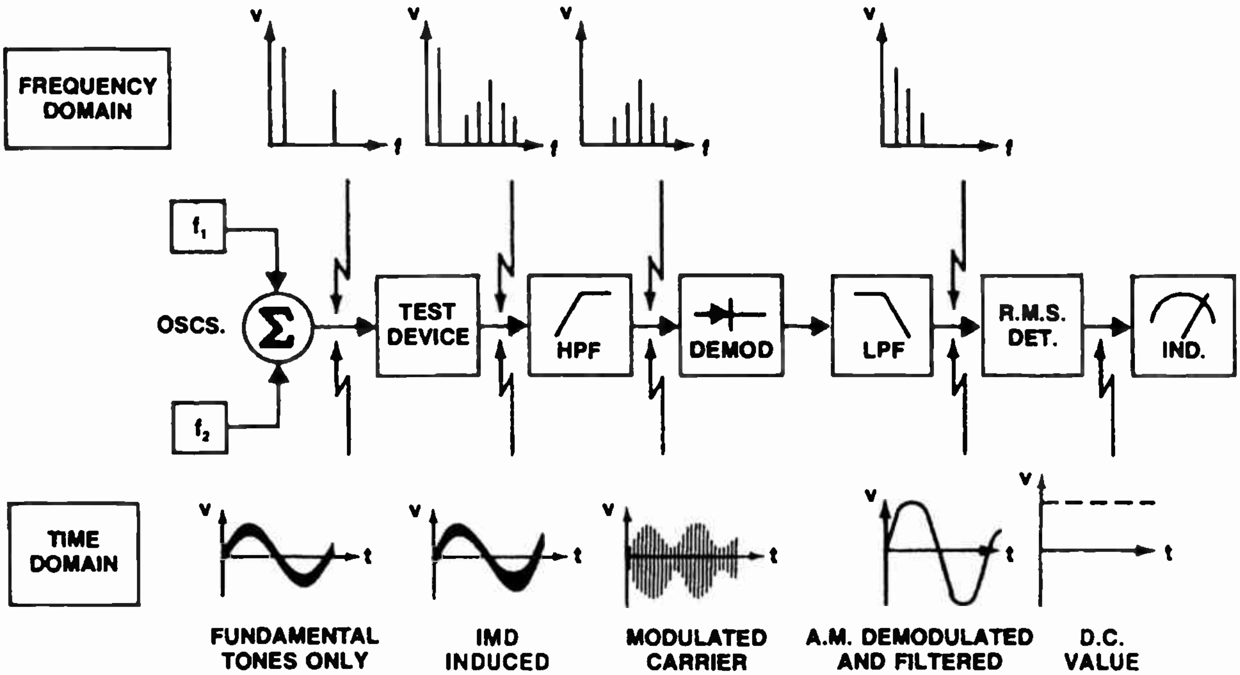


Figure 7.1-17. SMPTE IMD analyzer.

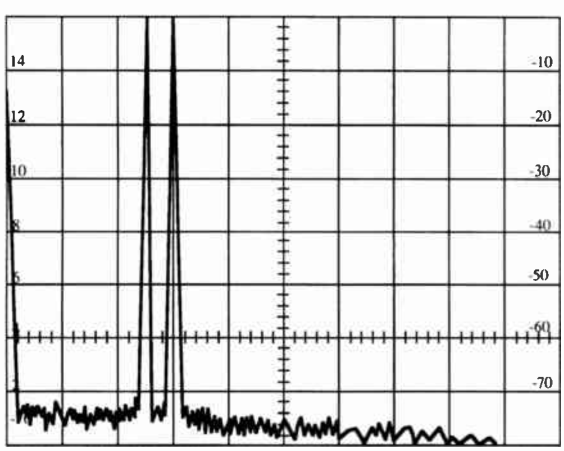


Figure 7.1-18. Spectrum of CCIF IMD test signal. (V: 10 dB/div; H: 2 kHz/div).

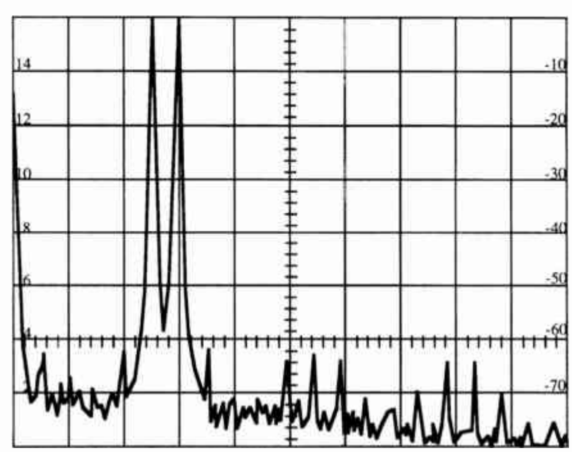


Figure 7.1-19. CCIF IMD measurement example. (V: 10 dB/div; H: 2 kHz/div).

order products, $2f_3 - f_4$ and $2f_4 - f_3$, to the sum of the amplitudes of the two driving frequencies, f_3 and f_4:

$$\text{CCIF IMD\%} = \left[\frac{A_{(2f_3-f_4)} + A_{(2f_4-f_3)}}{A_{f_3} + A_{f_4}}\right] \times 100 \quad (12)$$

(third order)

In the case of Figure 7.1-19, the driving frequencies f_3 and f_4 are at 5 kHz and 6 kHz, even-order product at 1 kHz, and third-order products at 4 kHz and 7 kHz, respectively. Distortion percentages for this example are calculated as follows:

$$A_{f_3} = A_{f_4} = 0\text{ dBm} = 775\text{ mV}$$

$$A_{(f_4-f_3)} = -64\text{ dBm} = 0.49\text{ mV}$$

$$\text{CCIF IMD\%} = \left[\frac{0.49}{775 + 775}\right] \times 100 = 0.032\%$$

(second order)

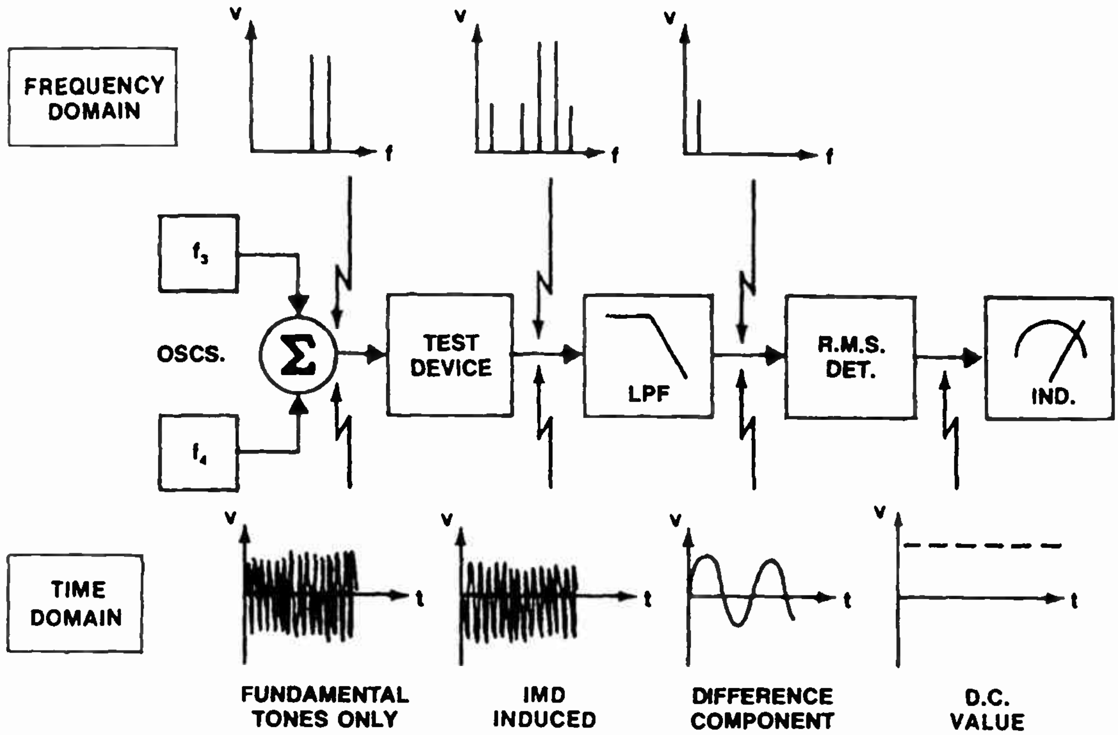


Figure 7.1-20. CCIF second-order IMD analyzer.

$$A_{(2f_3-f_4)} = -62 \text{ dBm} = 0.62 \text{ mV}$$

$$A_{(2f_4-f_3)} = -61 \text{ dBm} = 0.69 \text{ mV}$$

$$\text{CCIF IMD\%} = \left[\frac{0.62 + 0.69}{775 + 775}\right] \times 100 = 0.085\% \quad (13)$$

(third order)

It is a common practice for direct reading metered analyzers to measure only the amplitude of the difference product ($f_4 - f_3$) with respect to the driving signal amplitudes. The device that performs this task is called a *CCIF second-order difference frequency distortion analyzer.* Figure 7.1-20 illustrates how the measurement is made.

TIM distortion is found only in amplifiers that utilize negative feedback. When this feedback is excessive, a fast rising transient signal applied to the input of the amplifier can produce an internal overshoot that saturates the circuits in the amplifier.

The most popular procedure used to measure TIM distortion is the *sine square wave method.* The test signal employed uses a square wave (f_{sq}) to induce nonlinearity in the test device by saturating the amplifier's internal current, caused by its alternate rises and falls. Mixed with this square wave is a low level, high-frequency sine wave (f_{si}), which is unrelated harmonically. As defined, the frequency of the square wave is 3.18 kHz and that of the sine wave 15 kHz, where the peak-to-peak amplitude ratio of the former to the latter is four to one.[9] Before summation, the square wave is low-pass filtered using a first order design having a cutoff frequency of 30 kHz. This reduces the harmonics outside of the band of interest that could damage the device being tested. The composite wave form produced is shown in Figure 7.1-21 and spectrally in Figure 7.1-22.

Using the test setup shown in Figure 7.1-23, an amplifier can be measured for TIM. Mathematically, TIM distortion produced by the sine square wave method is defined as:

$$\text{TIM\%} = \frac{\sqrt{A1^2 + A2^2 + \cdots + A9^2}}{A_{si}} \times 100 \quad (14)$$

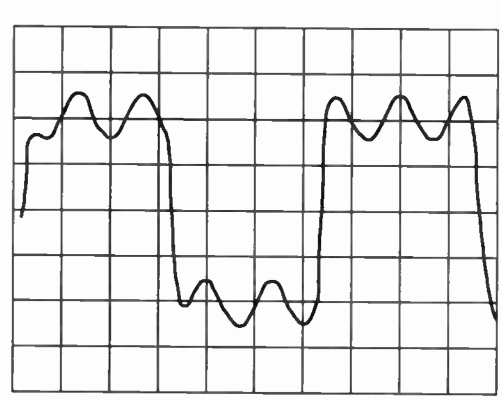


Figure 7.1-21. TIM test signal.

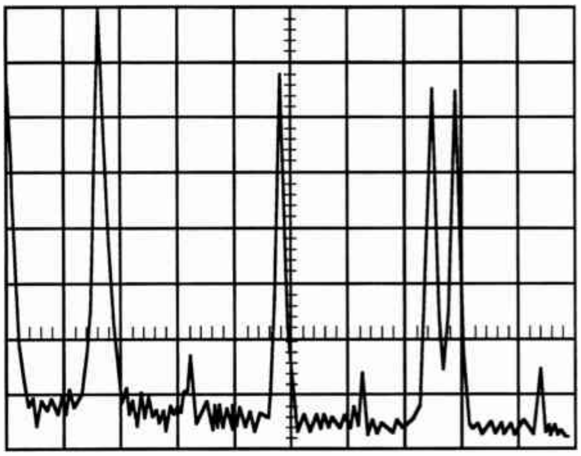


Figure 7.1-22. Spectral view of TIM test signal (V: 10 dB/div; H: 2 kHz/div).

where $A1$ through $A9$ represent the amplitudes of the distortion product present in the audio band and A_{si} is the amplitude of the sine wave. Values for the $A1$–$A9$ components are shown in Table 7.1-1.

Figure 7.1-24 shows the TIM distortion products produced by an amplifier that only displayed negligible THD, SMPTE IMD and CCIF IMD percentages. Using Formula 14, it can be determined that this amplifier is producing about 20% TIM distortion.

Other TIM measurement methods include a sawtooth wave method that takes amplifier slew rate into account and a noise-square wave method, where the sine wave of the sine-square wave method is replaced by a narrow band noise spectrum.[10]

Filtering as part of distortion measurement is often useful to remove components that are of little interest and as a diagnostic aid.[11] This is especially true with THD measurement. A 20 kHz or 30 kHz high-pass filter placed in series with the output of the device being measured is useful for testing broadcast equipment. This is often an acceptable practice, since the harmonics produced outside the transmission bandwidth can be eliminated, producing a more realistic result.

Table 7.1-1

Component	Relation	Frequency, kHz
A1	fsi – fsq	11.82
A2	fsi – 2fsq	8.64
A3	fsi – 3fsq	5.46
A4	fsi – 4fsq	2.28
A5	|fsi – 5fsq|	0.90
A6	|fsi – 6fsq|	4.08
A7	|fsi – 7fsq|	7.26
A8	|fsi – 8fsq|	10.44
A9	|fsi – 9fsq|	13.62

A high-pass filter also can serve as an important analytical aid. With a cutoff frequency in the 400 Hz range, it can be placed in series with the test device's output and used to determine the contribution of line frequency hum (60 Hz in the U.S.) to a THD measurement utilizing a fundamental frequency of 1 kHz or greater. Verification of adequate grounding used in the test setup also can be observed.

Care must be taken, however, when filtering is used in IM measurement. It is important to verify that in-band distortion components are not inadvertently removed. Also, extremely sharp cutoff filter designs may produce overshoot components that could affect measurement results.

An analyzer capable of performing many of the distortion tests previously described, as well as frequency response measurements, is shown in Figure 7.1-25.

NOISE MEASUREMENT

In audio engineering, noise is a random energy distribution in which individual spectral components are

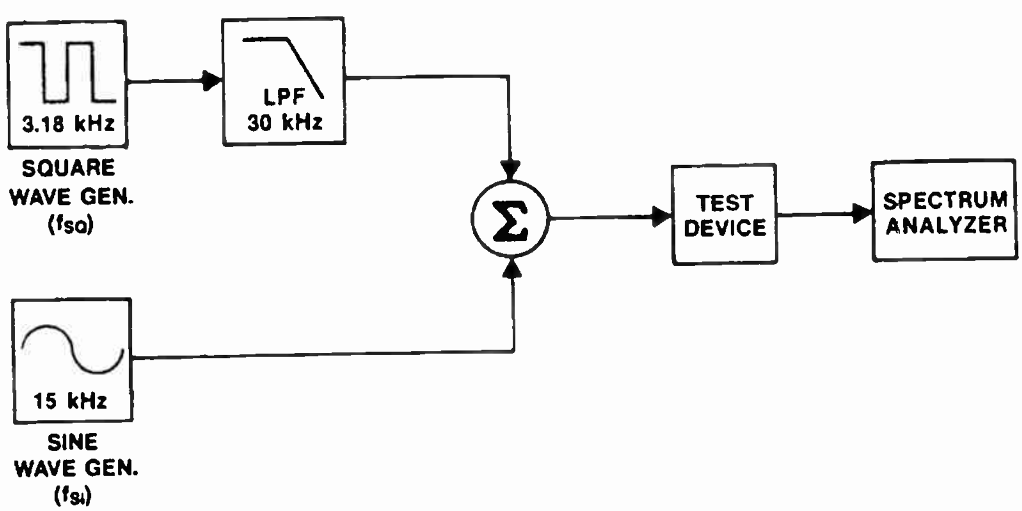


Figure 7.1-23. TIM distortion analysis test setup.

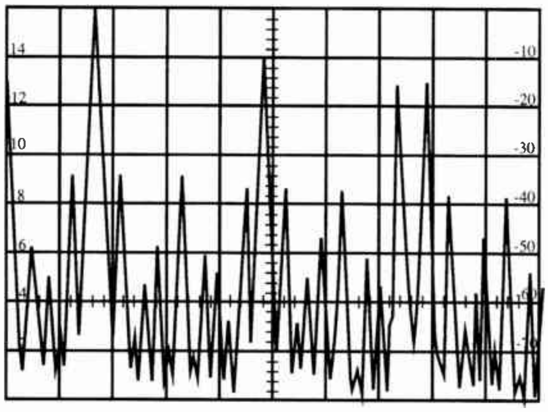


Figure 7.1-24. Spectral view of TIM distortion product measurement (V: dB/div; H: 2 kHz/div).

not clearly resolved. Primary sources of noise in analog circuits and amplifiers are in the resistive circuit elements.[12] It is important to control noise in amplifiers as their gain increases to preserve a high signal to noise ratio, which is the ratio of the operating signal level to the noise level inherent in the amplifier itself.

To understand the origin of the noise, we can model a passive resistive element as a noiseless resistor in series with a noise voltage generator, E_r:

$$E_r = \sqrt{4KTBR}\,(\text{volts}) \tag{15}$$

where:

K = Boltzman's constant (1.38×10^{-23} W-Sec/°K)
T = Temperature in degrees Kelvin
B = Noise bandwidth (Hz)
R = Resistance in Ω

As can be seen from the equation, noise voltage is a physical phenomenon that can be worsened by an increase in any of the variable factors. Therefore, noise cannot be eliminated but it can be reduced. This is often done by proper selection of the resistive components, because of an additional factor known as excess noise, which is proportional to the voltage drop across the resistor and related to the material from which it is made. Of the different available types, carbon composition resistors are prone to the most excess noise contribution while metal-film devices show the least.

At times, the actual spectral distribution of noise is of less importance than the noise voltage within a given bandwidth for comparison purposes. For audio frequencies, a 15 or 20 kHz bandwidth is of interest. With a low pass filter in this range connected in series with an amplifier output, and the input of the amplifier grounded, an unweighted but band limited noise measurement can be made. When the noise output level is obtained, it can be expressed as a ratio with a standard operating level and reference frequency. This produces an indication of the amplifier's SNR.

When the gain of the amplifier is known, this same technique can be used to determine equivalent input noise voltage. The voltage of the noise that would be found at the input of the amplifier if the amplifier were completely noiseless.[12]

The measurement of a noise voltage quantity over a given frequency bandwidth in order to determine SNR does not provide a complete characterization, because the noise spectrum can occupy all or part of the same bandwidth.[13] For example, two amplifiers with identical SNRs can sound very different because one may have a uniform noise spectrum and the other may have most of the noise concentrated over a limited frequency range. Hence, the latter amplifier would sound noisier than the former, having to do with the way the ear perceives the loudness of a signal that is uniform in amplitude across the radio band. To make comparative noise measurements more meaningful, several weighting filters have been used to alter noise spectra over the frequency band of interest.

"A" weighting is based on the inverse of early measurements by Fletcher and Munson of the ear's sensitivity at low sound pressure levels.[14] A more recently developed weighting curve utilizes the CCIR/ARM

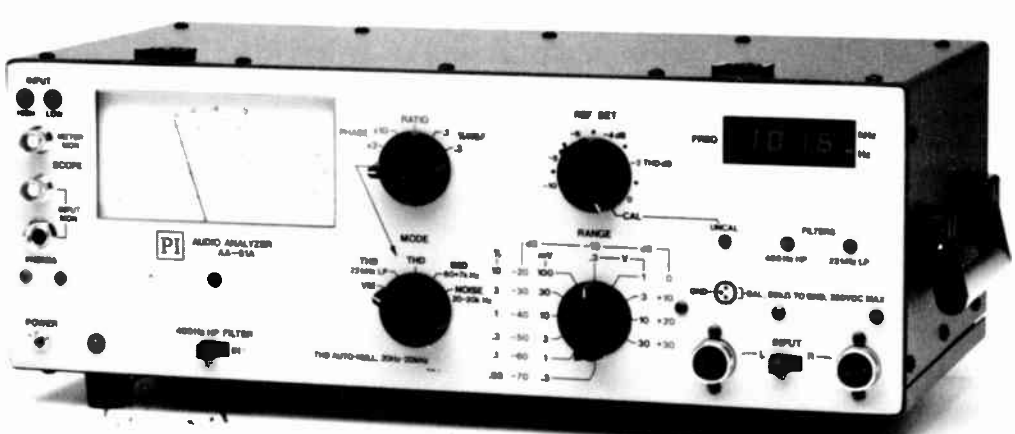


Figure 7.1-25. Analyzer capable of performing audio response and distortion measurements (Courtesy Potomac Instruments, Inc.).

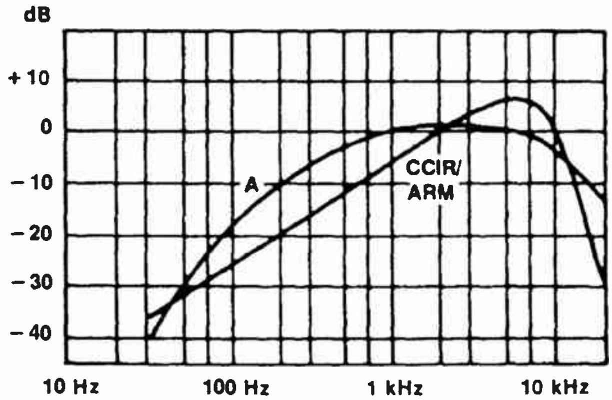


Figure 7.1-26. Noise measurement weighting schemes.

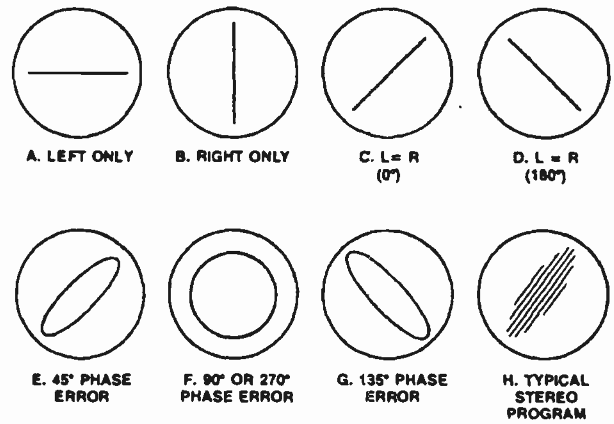


Figure 7.1-27. Interpretation of Lissajous patterns.

method, an updated scheme which places the zero dB reference at 2 kHz instead of 1 kHz.[15] It is believed that this method, which is based on the obtrusiveness as well as the levels of different kinds of noise, provides a more commercially acceptable result when used to characterize modern, wide range audio equipment. Figure 7.1-26 compares the two curves.

PHASE MONITORING AND MEASUREMENT

An (L + R) summation is the monophonic compatible signal for AM, FM, and TV stereo broadcasting. Separation information is usually transmitted via an (L − R) signal. Since these two signals are created through a summation and difference process of the original left and right channel stereophonic source, it is important that they can be recombined properly at the receiver.[16] Amplitude and phase errors must be minimized in the transmission system to accomplish this goal. Phase measurement is important in accomplishing this task.

In a stereo program system, if left and right audio information is correlated but delayed in phase, the error would not be evident on a stereophonic receiver. A monophonic signal, however, would be degraded because of inexact summation. This problem is common to analog audio tape recording. Periodic azimuth adjustment or *phasing* of the heads is often done. A test tape containing a high frequency tone is played while azimuth is adjusted to minimize the difference signal between the two channels.

A *phase meter* can be used to simplify this task. A simple version would take phase and amplitude variances into account simultaneously by functioning as a two-input subtractor. When both characteristics are identical in each of the channels, the output becomes zero. A meter that measures only phase information compares the zero-crossing times of the two input signals, and the resulting time difference is used to generate a dc voltage proportional to the phase difference.[17] Phase detectors operating in this manner often limit the input signals in order to remove all amplitude information.

More popular, however, is the *Lissajous figure method,* involving the use of an oscilloscope in the X-Y mode. The patterns produced are shown in Figure 7.1-27. An oscilloscope is connected such that the left channel audio causes an X-axis sweep and right channel audio produces a Y-axis sweep as shown in Figure 7.1-27(a) and 7.1-27(b) respectively. When each channel contains the same program material, the pattern of Figure 7.1-27(c) is produced. This is the L + R axis. If one of the channels is inverted, the pattern of Figure 7.1-27(d) becomes evident (this is often called the L − R axis). Program material that follows this axis is said to be inverted in polarity because no sum or L + R information is present.

During alignment of analog tape reproducing equipment, when discrete tones are used, the patterns of Figures 7.1-27(e), 7.1-27(f), and 7.1-27(g) are commonly seen when phase errors exist between the two channels. Stereo program material, in unprocessed form, generally modulates the L + R axis while simultaneously deviating in the L − R direction to a lesser amount, as shown in Figure 7.1-27(h).

REFERENCES

1. Graeme, J., *Applications of Operational Amplifiers,* McGraw-Hill, pp. 132–139, 202–203, 1973.
2. "A New Standard Volume Indicator and Reference Level," *Proceedings of the IRE,* pp. 1–17, January 1940.
3. Harry, D., "Audio Program Analysis," as presented to the National Association of Broadcasters, April 1985.
4. IEEE Standard: Recommended Practice for Audio Program Level Measurement, DOC. G-2.1.2/13, 1988.
5. Hewlett Packard Co. "Using a Narrow Band Analyzer for Characterizing Audio Products," AN192, Oct. 1975.

6. Salek, S., "The NAB Test CD—Use and Applications," as presented to the National Association of Broadcasters, April 1989.
7. National Association of Broadcasters, The NAB Broadcast and Audio System Test CD Instruction and Applications Booklet, 1988.
8. Crown International Inc., BDP-2 The Audio Micro Computer, equipment manual, September 1981.
9. Skritek, P. and H. Pichler, "Extended Application of T.I.M. Test Procedures," preprint no. 1557, 64th AES Convention, Nov. 1979.
10. Takahashi, S. and S. Tanaka, "A New Method of Measuring Transient Intermodulation Distortion: A Comparison with the Conventional Method," preprint no. 1539, 64th AES Convention, Nov. 1979.
11. Tektronix, Inc., AA501 Distortion Analyzer, equipment manual, February 1984.
12. National Semiconductor Corporation, *Audio/Radio Handbook,* pp. 2.3–2.10, 1980.
13. Hewlett Packard Co., "Spectrum Analysis . . . Noise Measurements," AN150-4, April 1974.
14. National Association of Broadcasters, Technical Standards, Section 5, pg. 23, 1990.
15. Dolby, R., D. Robinson and K. Gundry, "CCIR/ARM: A Practical Noise-Measurement Method," Journal of the Audio Engineering Society, vol. 27, pp. 149–157, March 1979.
16. Mendenhall, G., "The Composite Signal-Key to Quality FM Broadcasting," Broadcast Electronics.
17. Graeme, J., *Designing with Operational Amplifiers,* McGraw-Hill, pp. 76–77, 251–253, 1977.

7.2
VIDEO SIGNAL ANALYSIS

JERRY WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

As video technology steams full speed ahead into the digital domain, the shortcomings and degradations associated with analog technology that video engineers have come to accept and deal with—are rapidly disappearing. In their place, however, are new problems. Digital devices and systems bring their own unique mix of issues that must be addressed, including:

- **Quantization.** The quantization process, by design, discards information. It takes an analog waveform with infinite variability and blocks it into a collection of bits, the number of which is determined by the bit length of the system.
- **Concatenation.** Defined as the connection of elements end-to-end, concatenation for video and audio describes the effects of chaining compression and decompression systems.
- **Video Processing.** It is commonly assumed that as long as a video clip is manipulated in the digital domain, it will not be degraded. In a general sense this is true, however, certain operations will discard information that cannot be recreated downstream. Changes in sizing, adjustment of color hue and saturation and adjustment of luminance values are just some of the operations that can result in degradation of the signal unless proper precautions are taken. Something as simple as improper gamma setup on monitors can result in a host of problems as the signal meanders through the production process. Once picture information is discarded, it cannot be completely recreated.
- **Transmission.** In order for a digital video signal to be useful, it usually must be moved from one spot to another. This almost always involves codecs and a transmission medium. This medium may be coax, fiber or a radio frequency link. With any of these systems, degradations are possible—some are more vulnerable than others. An RF link usually has the greatest level of exposure to interfering signals. Coax, on the other hand, is basically closed to outside influences but has a finite cable length over which reliable communications can take place.

The important message here is that digital is not always perfect and that the need for test equipment and quality control does not disappear simply because a room full of analog boxes is replaced with a computer workstation. Furthermore, just because the picture looks good on a local monitor does not mean that it will look good (or at least look the same) at the end of a terrestrial or satellite link.

New test instruments are rising to the challenge posed by the new technologies being introduced to the video production process. As the equipment used by broadcasters and video professionals becomes more complex, the requirements for advanced, specialized maintenance tools also increases. These instruments range from simply go/no-go status indicators to automated test routines with pre-programmed pass/fail limits. Video quality control efforts must focus on the overall system, not just a particular island.

The attribute that makes a good test instrument is really quite straightforward: accurate measurement of the signal under test. The attributes important to the user, however, usually involve the following:

- Affordability
- Ease of use
- Performance

Depending upon the application, the order of these criteria may be inverted (performance, ease of use, affordability). Suffice it to say, however, that all elements of these specifications combine to translate into the user's definition of the ideal instrument.

The memory functions of the new breed of instruments provide important new capabilities, including archiving test setups and reference waveforms for ongoing projects and comparative tests. Hundreds of files typically can be saved for later use. With automatic measurement capabilities, even a novice technician can perform detail-oriented measurements quickly and accurately.

In the rush to embrace advanced, specific-purpose test instruments, it is easy to overlook the grandparents of all video test devices—the *waveform monitor* and *vectorscope*. Just because they are not new to the scene does not mean that they have outlived their usefulness.

The waveform monitor and vectorscope still fill valuable roles in the test and measurement world. Both, of course, have their roots in the general purpose *oscilloscope*. This heritage imparts some important benefits. The scope is the most universal of all instruments, combining the best abilities of the human user and the machine. Electronic instruments are well equipped to quickly and accurately measure a given amplitude, frequency, or phase difference—they perform calculation-based tasks with great speed. The human user, however, is far superior to any machine in interpreting and analyzing an image. The waveform monitor and vectorscope presents to the user—in an instant—wealth of information that allows rapid characterization and understanding of the signal under consideration.

Assessment of Color Reproduction

A number of factors may contribute to poor color rendition in a display system. To assess the effect of these factors, it is necessary to define system objectives, then establish a method of measuring departures from the objectives. Visual image display may be categorized as follows:

- **Spectral color reproduction.** The exact reproduction of the spectral power distributions of the original stimuli. Clearly, this is not possible in a video system with three primaries.
- **Exact color reproduction.** The exact reproduction of tristimulus values. The reproduction is then a metameric match to the original. Exact color reproduction will result in equality of appearance only if the viewing conditions for the picture and the original scene are identical. These conditions include the angular subtense of the picture, luminance and chromaticity of the surround and glare. In practice, exact color reproduction often cannot be achieved because of limitations on the maximum luminance that can be produced on a color monitor.
- **Colorimetric color reproduction.** A variant of exact color reproduction in which the tristimulus values are proportional to those in the original scene. In other words, the chromaticity coordinates are reproduced exactly, but the luminances all are reduced by a constant factor. Traditionally, color video systems have been designed and evaluated for colorimetric color reproduction. If the original and the reproduced reference whites have the same chromaticity, if the viewing conditions are the same, and if the system has an overall gamma of unity, colorimetric color reproduction is indeed a useful criterion. These conditions, however, often do not hold. In which case, colorimetric color reproduction is inadequate.
- **Equivalent color reproduction.** The reproduction of the original color appearance. This might be considered as the ultimate objective, but it cannot be achieved because of the limited luminance generated in a display system.
- **Corresponding color reproduction.** A compromise by which colors in the reproduction have the same appearance that colors in the original would have had if they had been illuminated to produce the same average luminance level and the same reference white chromaticity as that of the reproduction. For most purposes, corresponding color reproduction is the most suitable objective of a color video system.
- **Preferred color reproduction.** A departure from the preceding categories that recognizes the preferences of the viewer. It is sometimes argued that corresponding color reproduction is not the ultimate aim for some display systems, such as color television, and that it should be taken into account that people prefer some colors to be different from their actual appearance. For example, suntanned skin color is preferred to average real skin color, and the sky is preferred bluer and foliage greener than they really are.

Even if corresponding color reproduction is accepted as the target, some colors are more important than others. For example, flesh tones must be acceptable—not obviously reddish, greenish, purplish or otherwise incorrectly rendered. Likewise, the sky must be blue and the clouds white, within the viewer's range of acceptance. Similar conditions apply to other well-known colors of common experience.

Chromatic Adaptation and White Balance

With properly adjusted cameras and displays, whites and neutral grays are reproduced with the chromaticity of D65. Tests have shown that such whites (and grays) appear satisfactory in home viewing situations even if the ambient light is of quite different color temperature. Problems occur, however, when the white balance is slightly different from one camera to the next or when the scene shifts from studio to daylight or vice versa. In the first case, unwanted shifts of the displayed white occur, whereas in the other, no shift occurs even though the viewer subconsciously expects a shift.

By always reproducing a white surface with the same chromaticity, the system is mimicking the human visual system, which adapts so that white surfaces always appear the same, whatever the chromaticity of the illuminant (at least within the range of common light sources). The effect on other colors, however, is more complicated. In video cameras, the white balance adjustment usually is made by gain controls on the R, G and B channels. This is similar to the von Kries model of human chromatic adaptation, although the R, G and B primaries of the model are not the same as the video primaries. It is known that the von Kries model does not accurately account for the appearance of colors after chromatic adaptation, and so it follows that making simple gain changes in a video camera is not the ideal approach. Nevertheless, this approach seems to work well in practice, and the viewer does not object to the fact, for example, that the relative increase in the luminances of reddish objects in tungsten light is lost.

Overall Gamma Requirements

Colorimetric color reproduction requires that the overall gamma of the system—including the camera, display, and any gamma-adjusting electronics—be unity. This simple criterion is the one most often used in the design of a video color rendition system. However, the more sophisticated criterion of corresponding color reproduction takes into account the effect of the viewing conditions. In particular, several studies have shown that the luminance of the surround is important. For example, a dim surround requires a gamma of about 1.5 for optimum color reproduction.

Display Resolution and Pixel Format

The pixel represents the smallest resolvable element of a display. The size of the pixel varies from one type of display to another. In a monochrome CRT, pixel size is determined primarily by the following factors:

- Spot size of the electron beam (the current density distribution)
- Phosphor particle size
- Thickness of the phosphor layer

The term pixel was developed in the era of monochrome television, and the definition was—at that time—straightforward. With the advent of color-triad-based CRTs and solid-state display systems, the definition is not nearly so clear.

For a color CRT, a single triad of red, green and blue phosphor dots constitutes a single pixel. This definition assumes that the mechanical and electrical parameters of the CRT will permit each triad to be addressed without illuminating other elements on the face of the tube. Most display systems, however, will not meet this criterion. Depending on the design, a number of triads may constitute a single pixel in a CRT display. A more all-inclusive definition for the pixel is: the smallest spatial information element as seen by the viewer.[1]

Dot pitch is one of the principal mechanical criteria of a CRT that determines, to a large extent, the resolution of the display. Dot pitch is defined as the center-to-center distance between adjacent green phosphor dots of the red, green and blue triad.

The pixel format is the arrangement of pixels into horizontal rows and vertical columns. For example, an arrangement of 640 horizontal pixels by 480 vertical pixels results in a 640 × 480 pixel format. This description is not a resolution parameter in itself, simply the arrangement of pixel elements on the screen. Resolution is the measure of the ability to delineate picture detail; it is the smallest discernible and measurable detail in a visual presentation.[2]

Pixel density is a parameter closely related to resolution, stated in terms of pixels per linear distance. Pixel density specifies how closely the pixel elements are spaced on a given display. It follows that a display with a given pixel format will not provide the same pixel density, or resolution, on a large screen (such as 19 in. diagonal) as on a small screen (such as 12 in. diagonal).

Television lines is another term used to describe resolution. The term refers to the number of discernible lines on a standard test chart. As before, the specification of television lines is not, in itself, a description of display resolution. A 525-line display on a 17 in. monitor will appear to a viewer to have greater resolution than a 525-line display on a 30 in. monitor. Pixel density is the preferred resolution parameter.

Contrast Ratio

The purpose of a video display is to convey information by controlling the illumination of phosphor dots on a screen, or by controlling the reflectance or transmittance of a light source. The contrast ratio specifies the observable difference between a pixel that is switched on and one that is in its corresponding off state:

$$C_r = \frac{L_{on}}{L_{off}}$$

Where:

C_r = contrast ratio of the display
L_{on} = luminance of a pixel in the on state
L_{off} = luminance of a pixel in the off state

The area encompassed by the contrast ratio is an important parameter in assessing the performance of a display. Two contrast ratio divisions typically are specified:

- **Small area.** Comparison of the on and off states of a pixel sized area
- **Large area.** Comparison of the on and off states of a group of pixels

For most display applications, the small area contrast ratio is the more critical parameter.

Color Bar Test Patterns

Color bars are the most common pattern for testing encoders, decoders and other video devices. The test pattern typically contains several bars filled with primary and complementary colors. There are many variants, which differ in color sequence, orientation, saturation and intensity.

The standard color bar sequence is white, yellow, cyan, green, magenta, red, blue and black. This sequence can be produced in RGB format by a simple 3 bit counter. The typical specification of color bar levels (ITU-R Rec. BT.471-1, *Nomenclature and Description of Color Bar Signals*) is a set of four numbers separated by slashes or dots and giving RGB levels as a percentage of reference white in the following sequence:

white bar/black bar/max colored bars/min colored bars

For example, 100/0/100/25 means 100% R, G and B on the white bar; 0% R, G and B on the black bar; 100% maximum of R, G and B on colored bars; and 25% minimum of R, G and B on colored bars (see Figure 7.2-1).

Some color bar patterns merit special names. For example, 100/0/75/0 bars are often called EBU bars or 75% bars, and 100/0/100/25 bars are known as BBC bars. Nevertheless, the CCIR four number nomenclature remains the only reliable specification system to designate the exact levels for color bar test patterns.

The SMPTE variation is a matrix test pattern according to SMPTE document 1-1990 *Alignment Color Bar Test Signal for Television Picture Monitors*, that consists of 67% of the field containing seven (without black) 75% color bars, plus 8% of the field with the *new chroma set* bars blue/black/magenta/black/cyan/

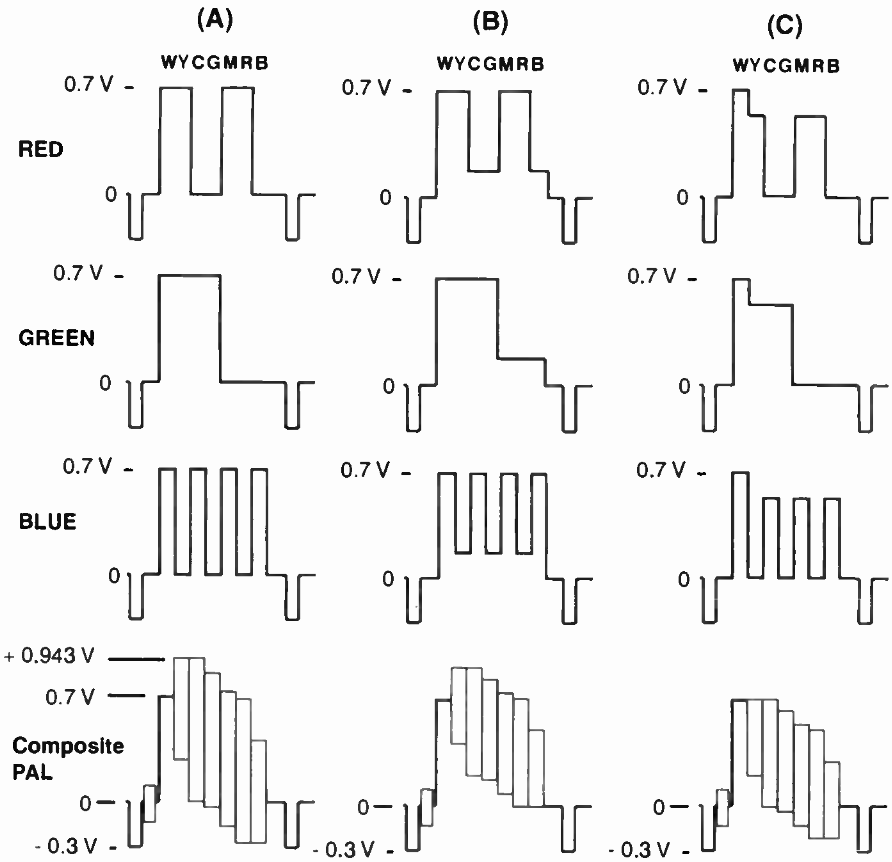


Figure 7.2-1. Examples of color bars nomenclature: (a) 100/0/100/0 color bars, (b) 100/0/100/25 color bars, (c) 100/0/75/0 color bars. *(Courtesy of Snell & Wilcox.)*

black/white) and the remaining 25% containing a combination of −I, white, Q, black and the black set signal (a version of PLUGE).

Chroma gain and chroma phase for picture monitors are usually adjusted by observing the standard encoded color bar signal with the red and green CRT guns switched off. The four visible blue bars are set for equal brightness. The use of the chroma set feature greatly increases the accuracy of this adjustment because it provides a signal with the blue bars to be matched vertically adjacent to each other. Because the bars are adjacent, the eye can easily perceive differences in brightness. This also eliminates effects resulting from shading or purity from one part of the screen to another.

The EIA color bar signal is a matrix test pattern according to RS-189-A, that consists of 75% of the field containing seven (without black) 75% color bars (same as SMPTE) and the remaining 25% containing a combination of −I, white, Q and black.

Conventional Video Measurements

Although there are a number of computer based television signal monitors capable of measuring video, sync, chroma and burst levels (as well as pulse widths and other timing factors), sometimes the best way to see what a signal is doing is to monitor it visually. The waveform monitor and vectorscope are oscilloscopes especially adapted for the video environment. The waveform monitor, like a traditional oscilloscope, operates in a voltage-versus-time mode. While an oscilloscope timebase can be set over a wide range of intervals, the waveform monitor timebase triggers automatically on sync pulses in the conventional TV signal, producing line and field rate sweeps, as well as multiple lines, multiple fields and shorter time intervals. Filters, clamps and other circuits process the video signal for specific monitoring needs. The vectorscope operates in an X-Y voltage-versus-voltage mode to display chrominance information. It decodes the signal in much the same way as a television receiver or a video monitor to extract color information and to display phase relationships. These two instruments serve separate, distinct purposes. Some models combine the functions of both types of monitors in one chassis with a single CRT display. Others include a communications link between two separate instruments.

Beyond basic signal monitoring, the waveform monitor and vectorscope provide a means to identify and analyze signal aberrations. If the signal is distorted, these instruments allow a technician to learn the extent of the problem and to locate the offending equipment.

Although designed for analog video measurement and quality control, the waveform monitor and vectorscope still serve valuable roles in the digital video facility.

Basic Waveform Measurements

Waveform monitors are used to evaluate the amplitude and timing of video signals and to show timing relationships between two or more signals. The familiar color bar pattern is the only signal required for these basic tests. Figure 7.2-2 shows a typical waveform display of color bars. It is important to realize that all color bars are not created equal. Some generators offer a choice of 75% or 100% amplitude bars. Sync, burst and setup amplitudes remain the same in the two color bar signals, but the peak-to-peak amplitudes of high frequency chrominance information and low frequency luminance levels change. The saturation of color, a function of chrominance and luminance amplitudes, remains constant at 100% in both modes. The 75% bar signal has 75% amplitude with 100% saturation. In 100% bars, amplitude and saturation are both 100%.

Chrominance amplitudes in 100% bars exceed the maximum amplitude that should be transmitted. Therefore, 75% amplitude color bars, with no chrominance information exceeding 100 IRE, are the standard amplitude bars for NTSC. In the 75% mode, a choice of 100 or 75 IRE white reference level may be offered. Figure 7.2-3 shows 75% amplitude bars with a 100 IRE white level. Either white level can be used to set levels, but operators must be aware of which signal has been selected. SMPTE bars have a white level of 75 IRE as well as a 100 IRE white flag.

The vertical response of a waveform monitor depends upon filters that process the signal in order to display certain components. The flat response mode displays all components of the signal. A chroma response filter removes luminance and displays only chrominance. The low-pass filter removes chrominance, leaving only low frequency luminance levels in the display. Some monitors include an IRE filter, designed to average-out high-level, fine-detail peaks on a monochrome video signal. The IRE filter aids the operator in setting brightness levels. The IRE response removes most, but not all, of the chrominance.

If the waveform monitor has a dual filter mode, the operator can observe luminance levels and overall amplitudes at the same time. The instrument switches between the flat and low-pass filters. The line select mode is another useful feature for monitoring live signals.

Sync Pulses

The duration and frequency of the sync pulses must be monitored closely for conventional video. Most waveform monitors include 0.5 ms or 1 ms per division magnification (MAG) modes, which can be used to verify H-sync width between 4.4 μs and 5.1 μs. The width is measured at the −4 IRE point. On waveform

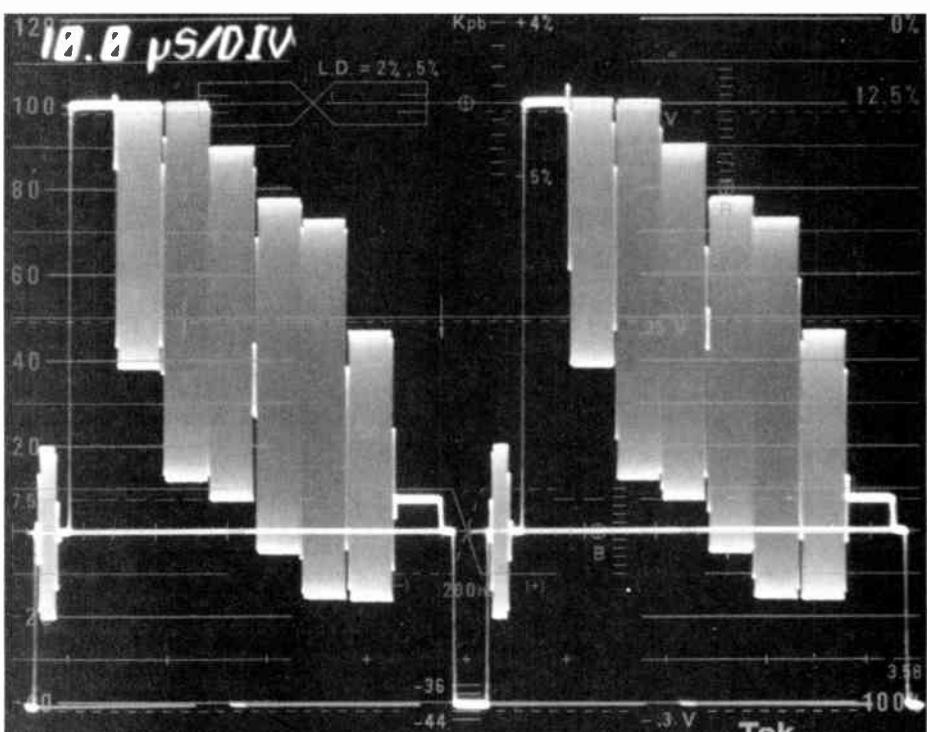


Figure 7.2-2. Waveform monitor display of a color-bar signal at the two-line rate. *(Courtesy of Tektronix.)*

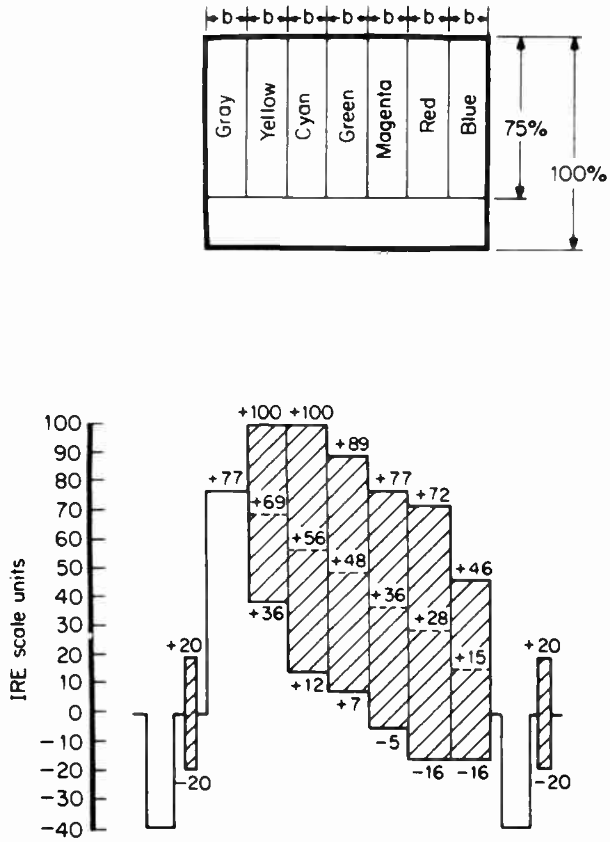


Figure 7.2-3. EIA RS189A color-bar displays: (a) color displays of gray and color bars, (b) waveform display of reference gray and primary/complementary colors, plus sync and burst.

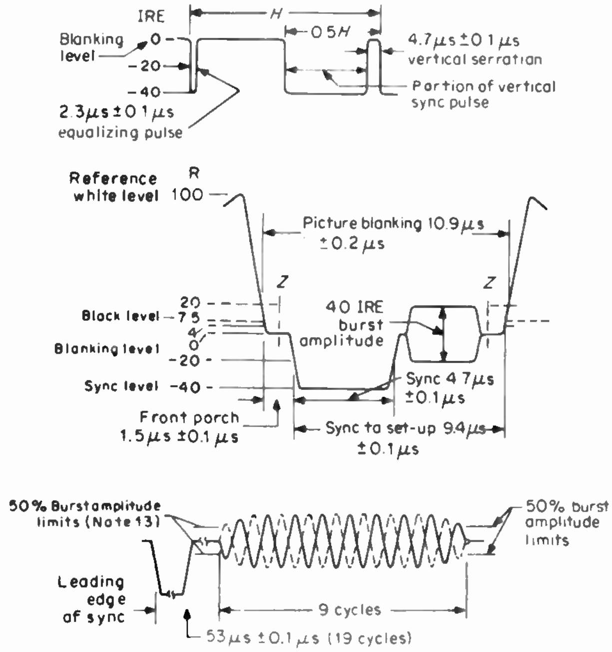


Figure 7.2-4. Sync pulse widths for NTSC.

monitors with good MAG registration, sync appearing in the middle of the screen in the 2 line mode remains centered when the sweep is magnified. Check the rise and fall times of sync and the widths of the front porch and entire blanking interval. Examine burst and verify there are between 8 and 11 cycles of subcarrier.

Check the vertical intervals for correct format, and measure the timing of the equalizing pulses and vertical sync pulses. The acceptable limits for these parameters are shown in Figure 7.2-4.

Basic Vectorscope Measurements

The vectorscope displays chrominance amplitudes, aids hue adjustments and simplifies the matching of burst phases of multiple signals. These functions require only the color bar test signal as an input stimulus. To evaluate and adjust chrominance in the TV signal, observe color bars on the vectorscope. The instrument should be in its calibrated gain position. Adjust the vectorscope phase control to place the burst vector at the 9 o'clock position. Note the vector dot positions with respect to the six boxes marked on the graticule. If everything is correctly adjusted, each dot will fall on the crosshairs of its corresponding box, as shown in Figure 7.2-5.

The chrominance amplitude of a video signal determines the intensity or brightness of color. If the amplitudes are correct, the color dots fall on the crosshairs in the corresponding graticule boxes. If vectors overshoot the boxes, chrominance amplitude is too large; if they undershoot, it is too small. The boxes at each color location can be used to quantify the error. In the radial direction, the small boxes indicate a ±2.5 IRE error from the standard amplitude. The large boxes indicate a ±20% error.

Figure 7.2-5. Vectorscope display of a color bar signal. *(Courtesy of Tektronix.)*

Other test signals, including a modulated staircase or multiburst, are used for more advanced tests. It is important to take a good look at how these signals appear immediately after they come out of the waveform generator. Knowing what the undistorted signal looks like simplifies the identification of distortions.

Line Select Features

Some waveform monitors and vectorscopes have line select capability. They can display one or two lines out of the entire video frame of 525 lines. (In the normal display, all of the lines are overlaid on top of one another.) The principal use of the single line feature is to monitor vertical interval test signals (VITS). VITS allows in-service testing of the transmission system. A typical VITS waveform is shown in Figure 7.2-6. A full field line selector drives a picture monitor output with an intensifying pulse. The pulse causes a single horizontal line on a picture monitor to be highlighted. This indicates where the line selector is within the frame to correlate the waveform monitor display with the picture.

Distortion Mechanisms

The video system should respond uniformly to signal components of different frequencies. In an analog system, this parameter is generally evaluated with a waveform monitor. Different signals are required to check the various parts of the frequency spectrum. If the signals are all faithfully reproduced on the waveform monitor screen after passing through the video system, it is safe to assume that there are no serious frequency response problems.

At very low frequencies, look for externally introduced distortions, such as power line hum or power supply ripple, and distortions resulting from inadequacies in the video processing equipment. Low frequency distortions usually appear on the video image as flickering or slowly varying brightness. Low frequency interference can be seen on a waveform monitor when the dc restorer is set to the slow mode and a 2 field sweep is selected. Sine wave distortion from ac power line hum may be observed in Figure 7.2-7. A bouncing APL signal can be used to detect distortion in the video chain. Vertical shifts in the blanking and sync levels indicate the possibility of low frequency distortion of the signal.

Field rate distortions appear as a difference in shading from the top to the bottom of the picture. A field rate 60 Hz square wave is best suited for measuring field rate distortion. Distortion of this type is observed as a tilt in the waveform in the 2 field mode with the dc restorer off.

Line rate distortions appear as streaking, shading or poor picture stability. To detect such errors, look for tilt in the bar portion of a pulse-and-bar signal. The waveform monitor should be in the 1H or 2H mode with the fast dc restorer selected for the measurement.

The multiburst signal is used to test the high frequency response of a system. The multiburst includes packets of discrete frequencies within the television passband, with the higher frequencies toward the right of each line. The highest frequency packet is at about 4.2

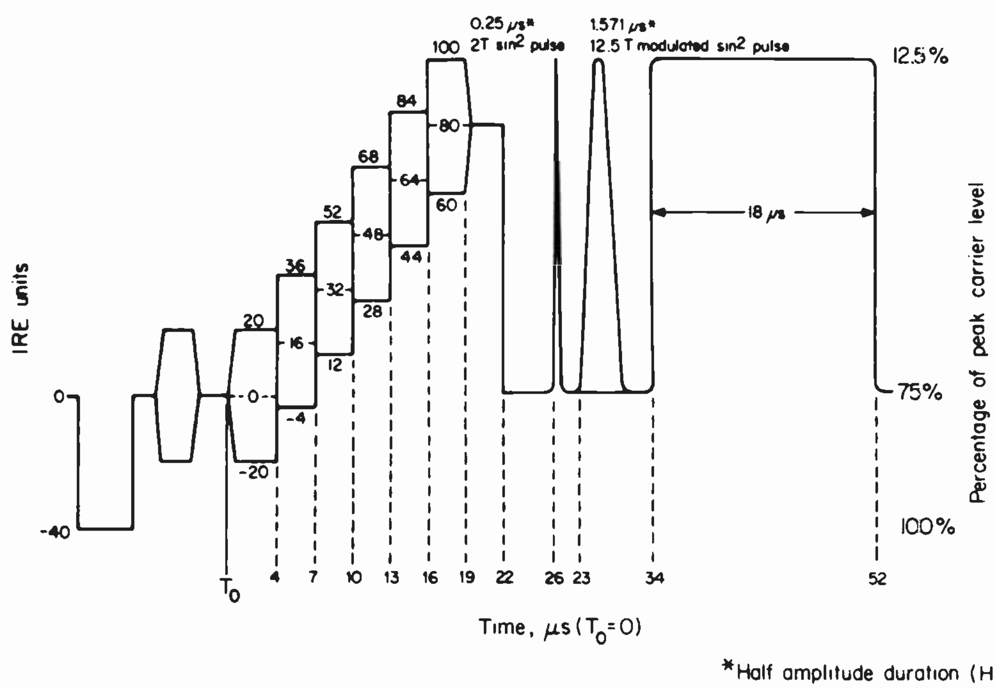


Figure 7.2-6. Composite vertical interval test signal (VITS) inserted in field 1, line 18. The video level in IRE units is shown on the left; the radiated carrier signal is shown on the right.

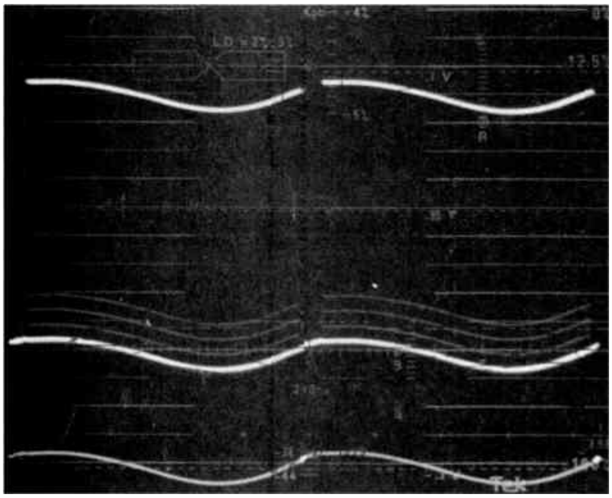


Figure 7.2-7. Waveform monitor display showing additive 60 Hz degradation. *(Courtesy of Tektronix.)*

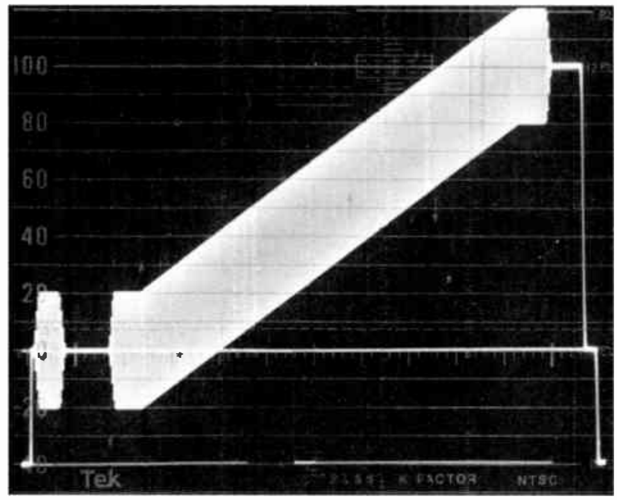


Figure 7.2-9. Waveform monitor display of a modulated ramp signal. *(Courtesy of Tektronix.)*

MHz, the upper frequency limit of the NTSC system. The next packet to the left is near the color subcarrier frequency (3.58 MHz) for checking the chrominance transfer characteristics. Other packets are included at intervals down to 500 kHz. The most common distortion is high frequency rolloff, seen on the waveform monitor as reduced amplitude packets at higher frequencies (see Figure 7.2-8). The television picture exhibits loss of fine detail and color intensity when such impairments are present. High frequency peaking, appearing on the waveform as higher amplitude packets at the higher frequencies, causes ghosting on the picture.

Differential Phase

Differential phase (df) distortion occurs if a change in luminance level produces a change in the chrominance phase. If the distortion is severe, the hue of an object will change as its brightness changes. A modulated staircase or ramp is used to quantify the problem. Either signal places chrominance of uniform amplitude and phase at different luminance levels. Figure 7.2-9 shows a 100 IRE modulated ramp. Because df can change with changes in APL, measurements at the center and at the two extremes of the APL range are necessary.

To measure df with a vectorscope, increase the gain control until the vector dot is on the edge of the graticule circle. Use the phase shifter to set the vector to the 9 o'clock position. Phase error appears as circumferential elongation of the dot. The vectorscope graticule has a scale marked with degrees of df error. Figure 7.2-10 shows a df error of 5°.

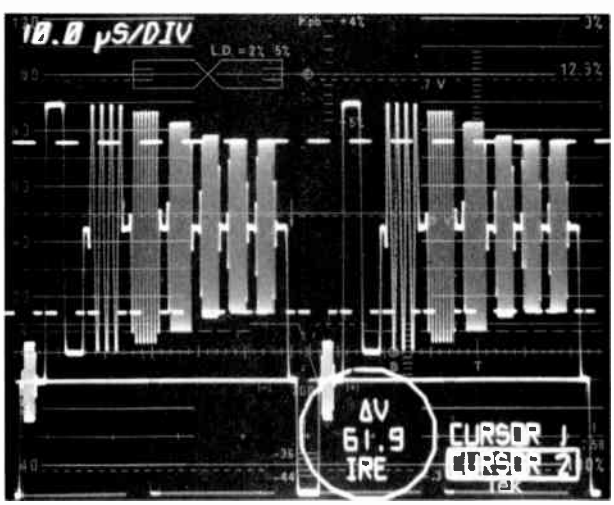


Figure 7.2-8. Waveform monitor display of a multiburst signal showing poor high frequency response. *(Courtesy of Tektronix.)*

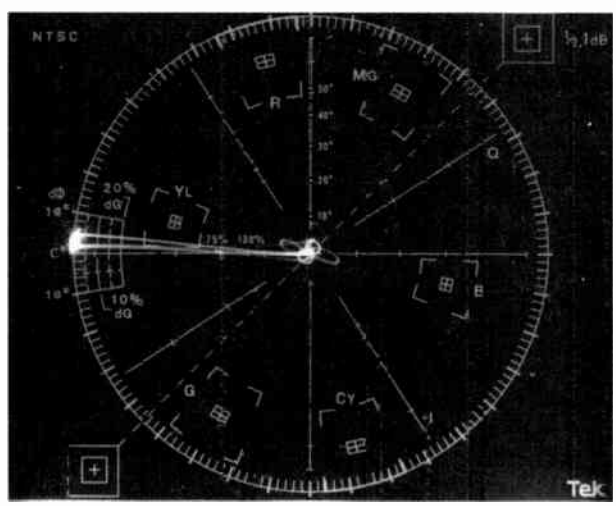


Figure 7.2-10. Vectorscope display showing 5° differential phase error. *(Courtesy of Tektronix.)*

More information can be obtained from a swept R-Y display, which is a common feature of waveform monitor and vectorscope systems. If one or two lines of demodulated video from the vectorscope are displayed on a waveform monitor, differential phase appears as tilt across the line. In this mode, the phase control can be adjusted to place the demodulated video on the baseline, which is equivalent in phase to the 9 o'clock position of the vectorscope. Figure 7.2-11 shows a df error of 5° with the amount of tilt measured against a vertical scale. This mode is useful in troubleshooting applications. By noting where along the line the tilt begins, it is possible to determine at what dc level the problem starts to occur. In addition, field rate sweeps enable the operator to look at df over the field.

A variation of the swept R-Y display may be available in some instruments for precise measurement of differential phase. Highly accurate measurements can be made with a vectorscope that includes a precision phase shifter and a double-trace mode. This method involves nulling the lowest part of the waveform with the phase shifter, then using a separate calibrated phase control to null the highest end of the waveform. A readout in tenths of a degree is possible.

Differential Gain

Differential gain (dG) distortion refers to a change in chrominance amplitude with changes in luminance level. The vividness of a colored object changes with variations in scene brightness. The modulated ramp or staircase is used to evaluate this impairment with the measurement taken on signals at different APL points.

To measure differential gain with a vectorscope, set the vector to the 9 o'clock position and use the variable gain control to bring it to the edge of the graticule circle. Differential gain error appears as a lengthening of the vector dot in the radial direction. The dG scale at the left side of the graticule can be used to quantify the error. Figure 7.2-12 shows a dG error of 10%.

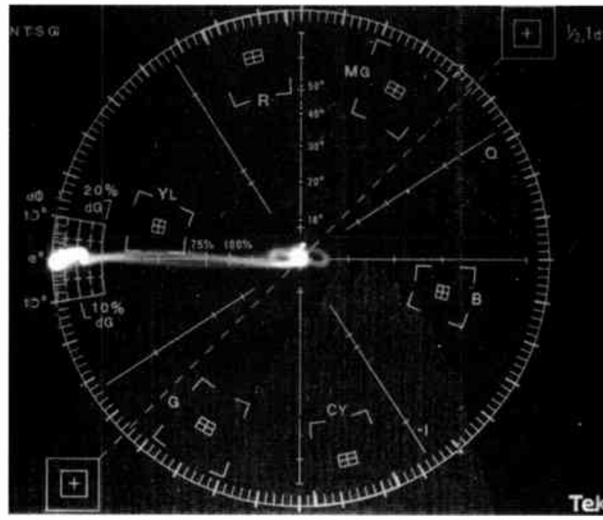


Figure 7.2-12. Vectorscope display of a 10% differential gain error. *(Courtesy of Tektronix.)*

Differential gain can be evaluated on a waveform monitor by using the chroma filter and examining the amplitude of the chrominance from a modulated staircase or ramp. With the waveform monitor in 1H sweep, use the variable gain to set the amplitude of the chrominance to 100 IRE. If the chrominance amplitude is not uniform across the line, there is dG error. With the gain normalized to 100 IRE, the error can be expressed as a percentage. Finally, dG can be precisely evaluated with a swept display of demodulated video. This is similar to the single trace R-Y methods for differential phase. The B-Y signal is examined for tilt when the phase is set so that the B-Y signal is at its maximum amplitude. The tilt can be quantified against a vertical scale.

Figure 7.2-11. Display showing a differential phase error of 5.95° as a tilt on the vertical scale. *(Courtesy of Tektronix.)*

Automated Video Signal Measurement

Video test instruments based on microcomputer systems provide the maintenance engineer with the ability to rapidly measure a number of parameters with exceptional accuracy. Automated instruments offer a number of benefits, including reduced setup time, test repeatability, waveform storage and transmission capability and remote control of instrument/measurement functions. Typical features of this class of instrument include the following:

- Waveform monitor functions
- Vectorscope monitor functions
- Picture display capability
- Automatic analysis of an input signals
- RS-232 and/or GPIB I/O ports

Figure 7.2-13 shows a block diagram of a representative automated video test instrument. Sample output waveforms are shown in Figure 7.2-14.

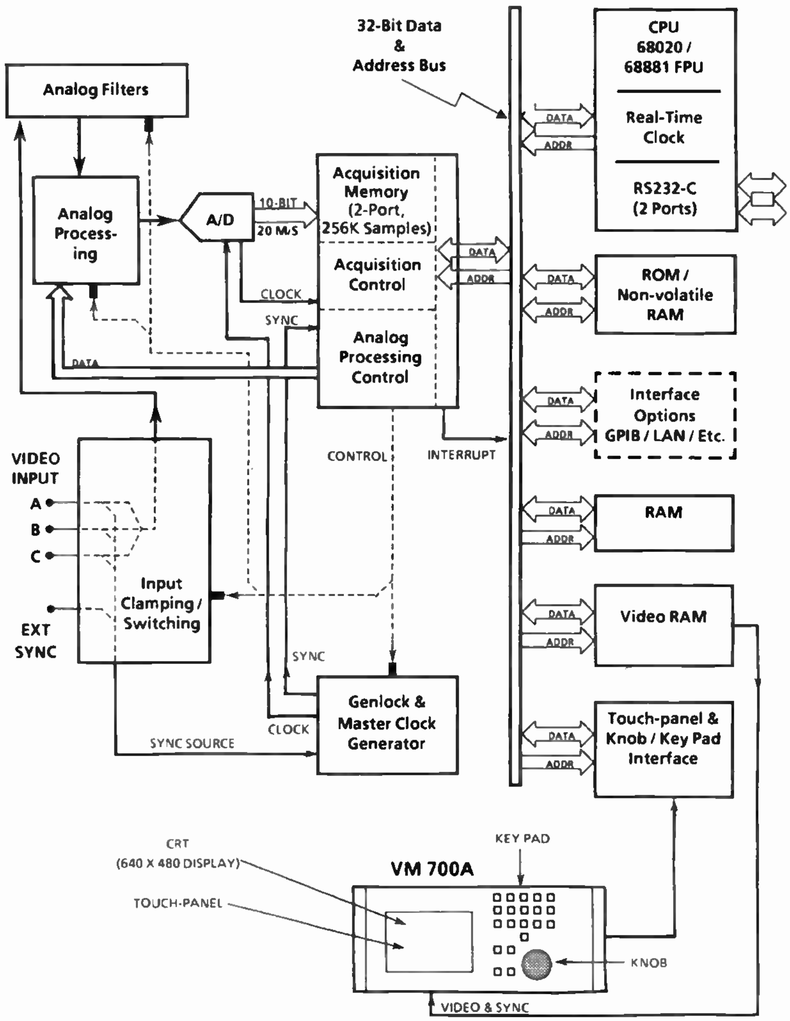
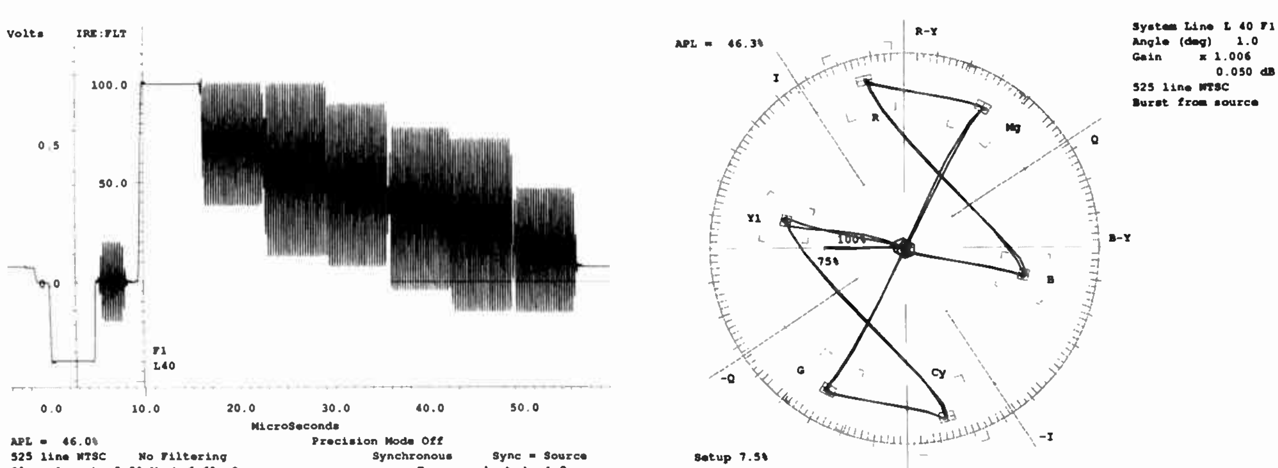


Figure 7.2-13. Block diagram of an automated video test instrument. *(Courtesy of Tektronix.)*



(a) (b)

Figure 7.2-14. Automated video test instrument output charts: (a) waveform monitor mode display of color bars, (b) vectorscope mode display of color bars. *(Courtesy of Tektronix.)*

Automated test instruments offer the end-user the ability to observe signal parameters and to make detailed measurements. The instrument depicted in Figure 7.2-13 offers a Measure Mode, in which a captured waveform can be expanded and individual elements of the waveform examined. This instrument provides interactive control of measurement parameters, as well as graphical displays and digital readouts of the measurement results. Figure 7.2-15 illustrates measurement of sync parameters on an NTSC waveform.

Another feature of many automated video measurement instruments is the ability to set operational limits and parameters that—if exceeded—are brought to the attention of an operator. In this way, operator involvement is required only if the monitored signal varies from certain preset limits. Figure 7.2-16 shows an example error log.

				Violated Limits Lower	Upper	
Bar Top	——	% Carr	**	10.0	15.0	Bar Not Found
Blanking Level	——	% Carr	**	72.5	77.5	ZC Pulse Unselected
Bar Amplitude	——	IRE	**	96.0	104.0	Bar Not Found
Sync Variation	——	% Carr	**	0.0	5.0	ZC Pulse Unselected
VIRS Setup	——	% Bar	**	5.0	10.0	Not Found
VIRS Luminance Ref	——	% Bar	**	45.0	55.0	Not Found
VIRS Chroma Ampl	——	% Burst	**	90.0	110.0	Not Found
VIRS Chroma Ampl	——	% Bar	**	36.0	44.0	Not Found
VIRS Chroma Phase	——	Deg	**	-10.0	10.0	Not Found
Line Time Distortion	——	%	**	0.0	2.0	No Composite VITS
Pulse/Bar Ratio	——	%	**	94.0	106.0	No Composite VITS
2T Pulse K-Factor	——	% Kf	**	0.0	2.5	No Composite VITS
IEEE-511 ST Dist	——	% SD	**	0.0	3.0	No Composite VITS

Figure 7.2-16. Example error log for a monitored signal. *(Courtesy of Tektronix.)*

Applications of the Zone Plate Signal

The increased information content of advanced high-definition display systems requires sophisticated processing to make recording and transmission practical.[3] This processing uses various forms of bandwidth compression, scan-rate changes, motion-detection and motion-compensation algorithms and other techniques. Zone plate patterns are well suited to exercising a complex video system in the three dimensions of its signal spectrum: horizontal, vertical and temporal. Zone plate signals, unlike most conventional test signals, can be complex and dynamic. As a result, they are capable of simulating much of the detail and movement of actual video, exercising the system under test with signals representative of the intended application. These digitally generated and controlled signals also have other important characteristics needed in test waveforms for video systems.

A signal intended for meaningful testing of a video system must be carefully controlled, so that any departure from a known parameter of the signal is attributable to a distortion or other change in the system under test. The test signal also must be predictable, so that it can be accurately reproduced at other times or places. These constraints usually have led to test signals that are electronically generated. In a few special cases, a standardized picture has been televised by a camera or monoscope—usually for a subjective, but more detailed, evaluation of overall performance of the video system.

A *zone plate* is a physical optical pattern, which was first used by televising it in this way. Now that electronic generators are capable of producing similar patterns, the label zone plate is applied to the wide variety of patterns created by video test instruments.

Conventional test signals, for the most part limited by the practical considerations of electronic generation, have represented relatively simple images. Each signal is capable of testing a narrow range of possible distortions; several test signals are needed for a more complete evaluation. Even with several signals, this method may not reveal all possible distortions or allow study of all pertinent characteristics. This is true especially in video systems employing new forms of sophisticated signal processing.

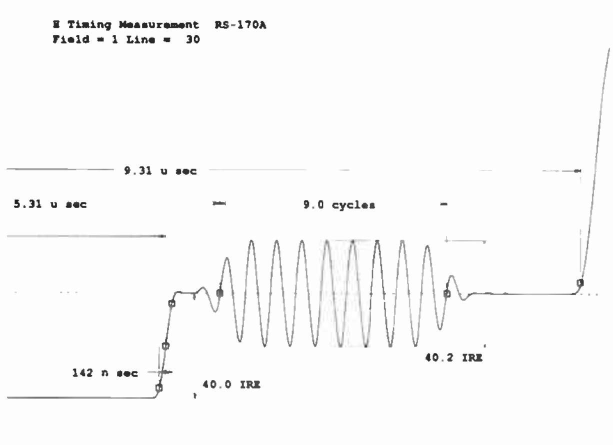
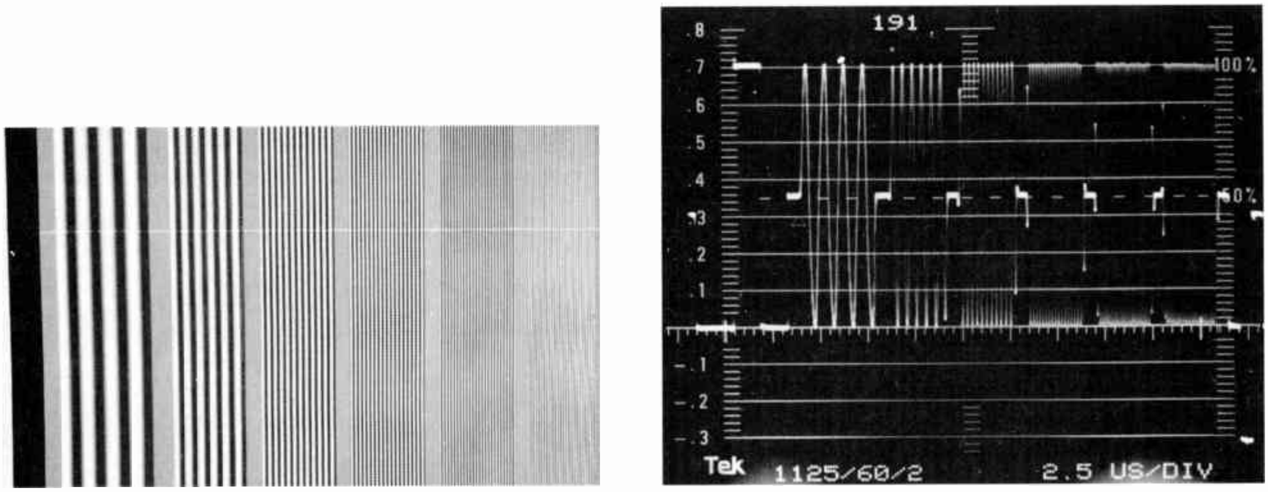


Figure 7.2-15. Expanded view of a sync waveform with measured parameters. *(Courtesy of Tektronix.)*

Simple Zone Plate Patterns

The basic testing of a video communication channel historically has involved the application of several single frequencies—in effect, spot-checking the spectrum of interest.[3] A well-known and quite practical adaptation of this idea is the *multiburst signal*, shown in Figure 7.2-17. This test waveform has been in use since the earliest days of video. The multiburst signal provides several discrete frequencies along a TV line. Figure 7.2-17(a) and other photographs in this section, show the beat effects introduced by the screening process used for photographic printing. This is largely unavoidable. The screening process is quite similar to the scanning or sampling of a television image—the patterns are designed to identify this type of problem.

The frequency-sweep signal is an improvement on multiburst. Although harder to implement in earlier generators, it was easier to use. The frequency-sweep signal, illustrated in Figure 7.2-18, varies the applied



(a) (b)

Figure 7.2-17. Multiburst video test waveform: (a, left) picture display, (b, right) multiburst signal as viewed on a waveform monitor (1H). *(Courtesy of Tektronix.)*

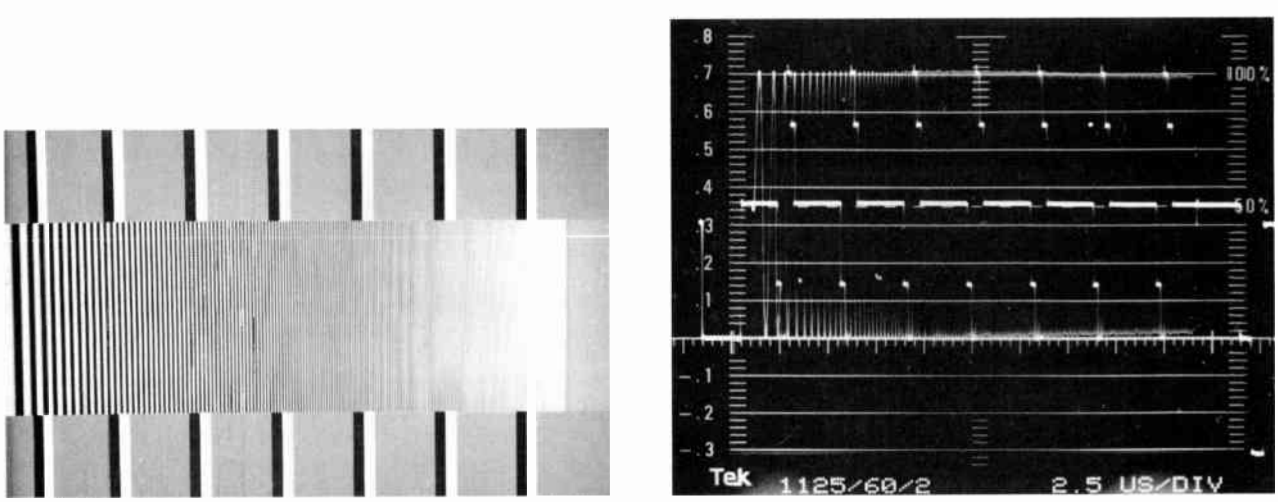
signal frequency continuously along the TV line. In some cases, the signal is swept as a function of the vertical position (field time). Even in these cases, the signal being swept is appropriate for testing the spectrum of the horizontal dimension of the picture.

Figure 7.2-19 shows the output of a zone plate generator configured to produce a horizontal single frequency output. Figure 7.2-20 shows a zone plate generator configured to produce a frequency-sweep signal. Electronic test patterns, such as these, may be used to evaluate the following system characteristics:

- Channel frequency response
- Horizontal resolution
- Moiré effects in recorders and displays
- Other impairments

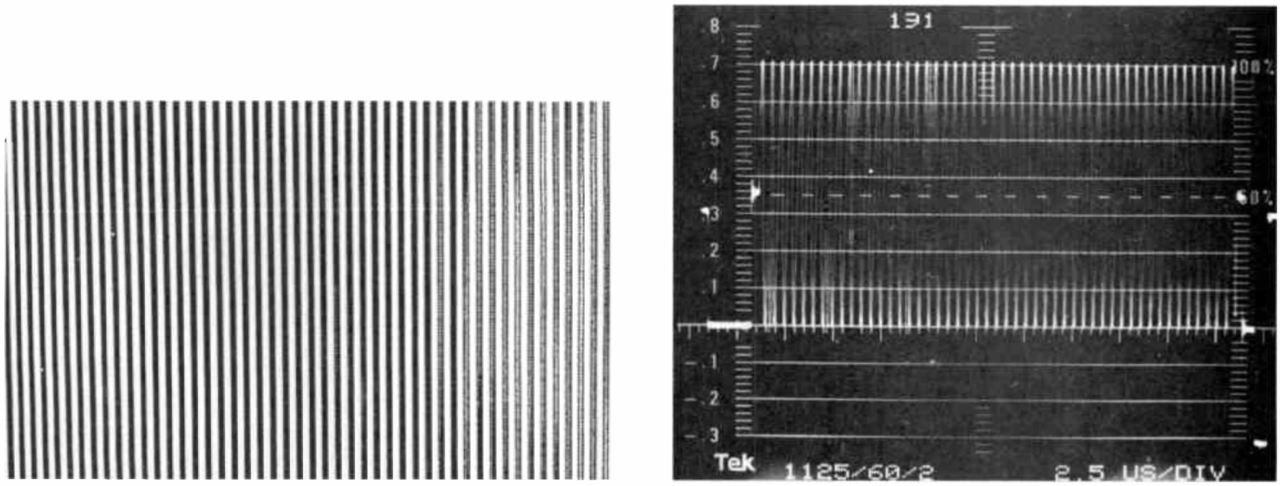
Traditionally, patterns that test vertical (field) response have been less frequently used. As new technologies implement conversion from interlaced to progressive scan, line-doubling display techniques, vertical antialiasing filters, scan conversion, motion detection or other processes that combine information from line-to-line, vertical testing patterns will be more in demand.

In the vertical dimension, as well as the horizontal, tests may be done at a single frequency or with a frequency-sweep signal. Figure 7.2-21 illustrates a magnified vertical-rate waveform display. Each dash in the photo represents one horizontal scan line. Sampling of vertical frequencies is inherent in the scanning process, and the photo shows the effects on the signal waveform. Note also that the signal voltage remains



(a) (b)

Figure 7.2-18. Conventional sweep-frequency test waveform: (a, left) picture display, (b, right) waveform monitor display, with markers (1H). *(Courtesy of Tektronix.)*



(a) (b)

Figure 7.2-19. Single horizontal frequency test signal from a zone plate generator: (a, left) picture display, (b, right) waveform monitor display (1H). *(Courtesy of Tektronix.)*

constant during each line, changing only from line-to-line in accord with the vertical dimension sine function of the signal. Figure 7.2-22 shows a vertical frequency-sweep picture display.

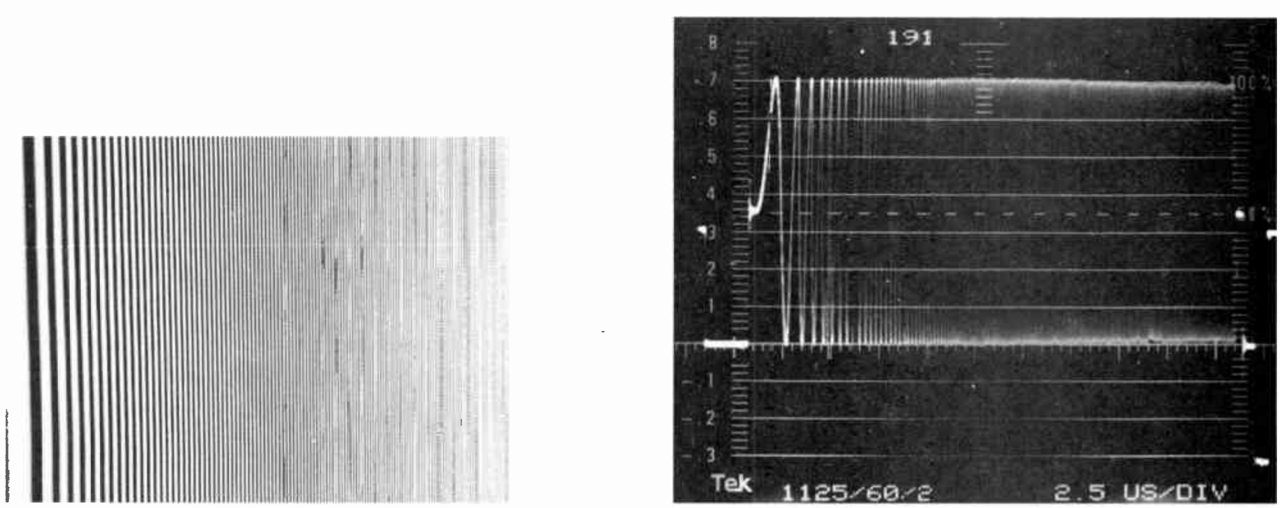
The horizontal and vertical sine waves and sweeps are quite useful, but they do not use the full potential of a zone plate signal source.

Producing the Zone Plate Signal

A zone plate signal is created in real time by a test signal generator.[3] The value of the signal at any instant is represented by a number in the digital hardware. This number is incremented as the scan progresses through the three dimensions that define a point in the video image: horizontal position, vertical position and time.

The exact method by which these dimensions alter the number is controlled by a set of coefficients. These coefficients determine the initial value of the number and control the size of the increments as the scan progresses along each horizontal line, from line-to-line vertically and from field-to-field temporally. A set of coefficients uniquely determines a pattern, or a sequence of patterns, when the time dimension is active.

This process produces a sawtooth waveform; overflow in the accumulator holding the signal number effectively resets the value to zero at the end of each cycle of the waveform. Usually, it is desirable to minimize the harmonic energy content of the output signal; in this case, the actual output is a sine function of the number generated by the incrementing process.



(a) (b)

Figure 7.2-20. Horizontal frequency-sweep test signal from a zone plate generator: (a, left) picture display, (b, right) waveform monitor display (1H). *(Courtesy of Tektronix.)*

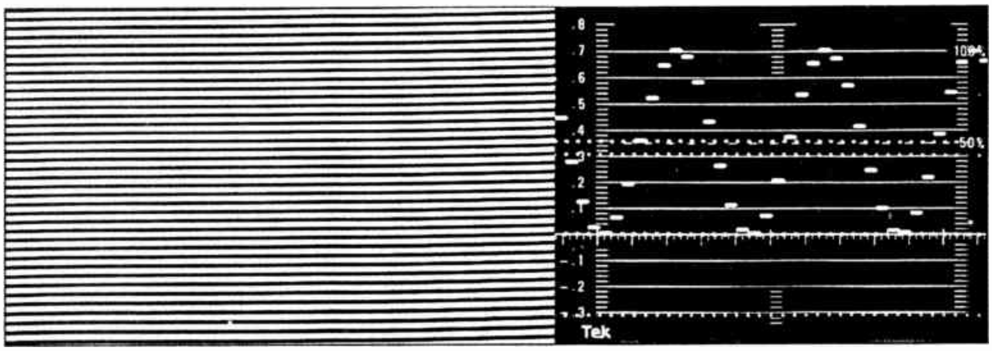


Figure 7.2-21. Single vertical frequency test signal: (a, left) picture display, (b, right) magnified vertical-rate waveform, showing the effects of scan sampling. *(Courtesy of Tektronix.)*

Complex Patterns

A pattern of sine waves or sweeps in multiple dimensions may be produced, using the unique architecture of the zone plate generator.[3] The pattern shown in Figure 7.2-23, for example, is a single signal sweeping both horizontally and vertically. Figure 7.2-24 shows the waveform of a single selected line (line 263 in the 1125/60/2 HDTV system). Note that the horizontal waveform is identical to the one shown in Figure 7.2-18(b), even though the vertical dimension sweep is now also active. Actually, different lines will give slightly different waveforms. The horizontal frequency and sweep characteristics will be identical, but the starting phase must be different from line-to-line to construct the vertical signal.

Figure 7.2-25 shows a 2-axis sweep pattern that is most often identified with zone plate generators, perhaps because it quite closely resembles the original optical pattern. In this circle pattern, both horizontal and vertical frequencies start high, sweep to zero (in the center of the screen), and sweep up again to the end of their respective scans. The concept of 2-axis sweeps actually is more powerful than it might first appear. In addition to purely horizontal or vertical effects, there are possible distortions or artifacts that are apparent only with simultaneous excitation in both axes. In other words, the response of a system to diagonal detail may not be predictable from information taken from the (individual) horizontal and vertical responses.

The Time (Motion) Dimension

Incrementing the number in the accumulator of the zone plate generator from frame-to-frame (or field-to-field in an interlaced system) creates a predictably different pattern for each vertical scan.[3] This, in turn, creates apparent motion and exercises the signal spectrum in the temporal dimension. Analogous to the single frequency and frequency-sweep examples given previously, appropriate setting of the time-related coefficients will create constant motion or motion sweep (acceleration).

Specific motion-detection and interpolation algorithms in a system under test may be exercised by determining the coefficients of a critical sequence of patterns. These patterns then may be saved for subsequent testing during development or adjustment. In an operational environment, appropriate response to a critical sequence could ensure expected operation of the equipment or facilitate fault detection.

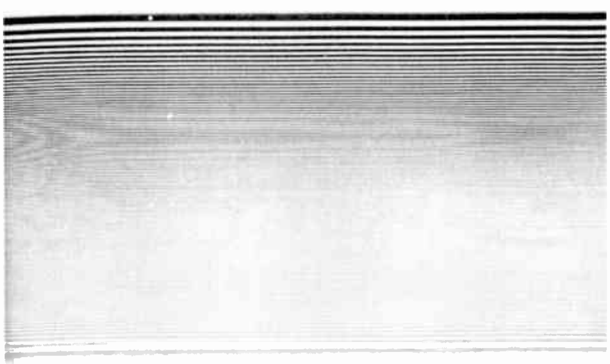


Figure 7.2-22. Vertical frequency-sweep picture display. *(Courtesy of Tektronix.)*

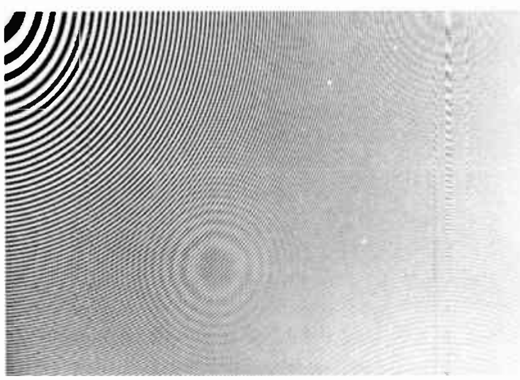


Figure 7.2-23. Combined horizontal and vertical frequency-sweep picture display. *(Courtesy of Tektronix.)*

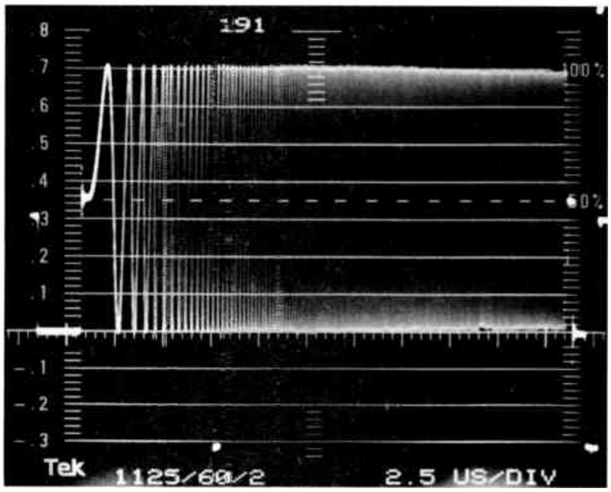


Figure 7.2-24. Combined horizontal and vertical frequency sweeps, selected line waveform display (1H). This figure shows the maintenance of horizontal structure in the presence of vertical sweep. *(Courtesy of Tektronix.)*

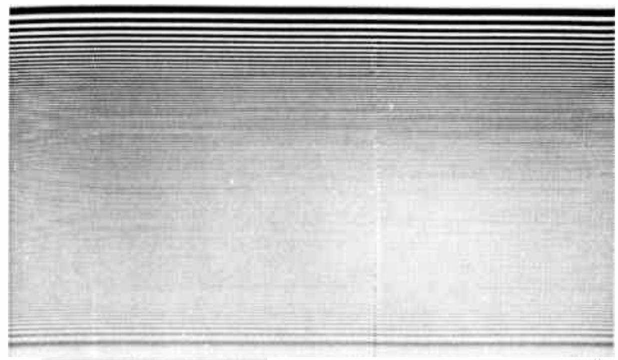


Figure 7.2-26. Vertical frequency-sweep picture display. *(Courtesy of Tektronix.)*

Although motion artifacts are difficult to portray in the still-image constraints of a printed book, the following example gives some idea of the potential of a versatile generator. In Figure 7.2-26, the vertical sweep maximum frequency has been increased to the point where it is zero-beating with the scan at the bottom of the screen. (The cycles/ph of the pattern matches the lines/ph per field of the scan.) Actually, in direct viewing, there is another noticeable artifact in the vertical center of the screen: a harmonic beat related to the gamma of the display CRT. Because of interlace, this beat flickers at the field rate. The photograph integrates the interfield flicker, thereby hiding the artifact, which is readily apparent when viewed in real time.

Figure 7.2-27 is identical to the previous photo, except for one important difference-upward motion of 1/2-cycle per field has been added to the pattern. Now the sweep pattern itself is integrated out, as is the first-order beat at the bottom. The harmonic effects in center screen no longer flicker, because the change of scan vertical position from field-to-field is compensated by a change in position of the image. The resulting beat pattern does not flicker and is easily photographed or, perhaps, scanned to determine depth of modulation.

A change in coefficients produces hyperbolic, rather than circular 2-axis patterns, as shown in Figure 7.2-28. Another interesting pattern, which has been used for checking complex codecs, is shown in Figure 7.2.29. This is also a moving pattern, which was altered slightly to freeze some aspects of the movement for the purpose of taking the photograph.

Display Measurement Techniques

A number of different techniques have evolved for measuring the static performance of picture display devices and systems.[4] Most express the measured device performance in a unique figure of merit or metric. Although each approach provides useful information, the lack of standardization in measurement techniques

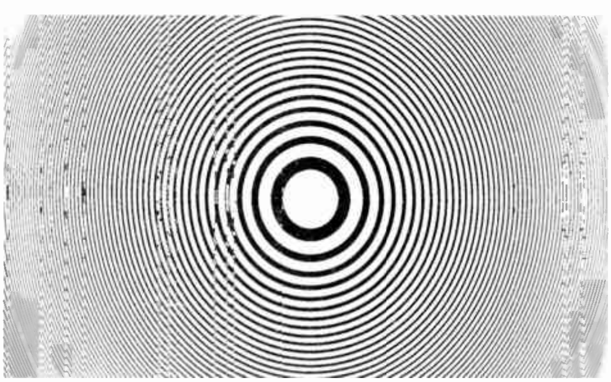


Figure 7.2-25. The best known zone plate pattern, combined horizontal and vertical frequency-sweeps with zero frequency in the center screen. *(Courtesy of Tektronix.)*

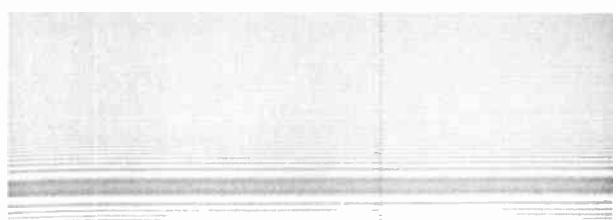


Figure 7.2-27. The same vertical sweep as shown in Figure 7.2-26, except that appropriate pattern motion has been added to freeze the beat pattern in the center screen for photography or other analysis. *(Courtesy of Tektronix.)*

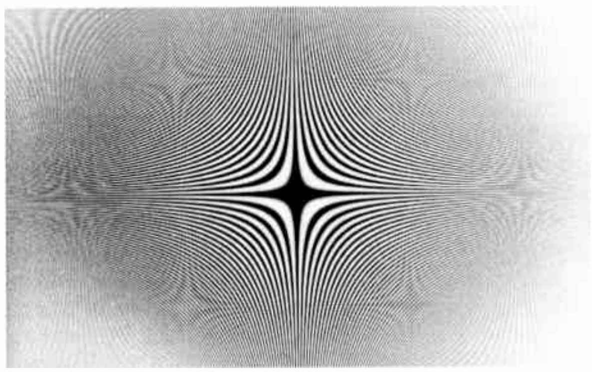


Figure 7.2-28. A hyperbolic variation of the 2-axis zone plate frequency-sweep. *(Courtesy of Tektronix.)*

makes it difficult or even impossible to directly compare the performance of a given class of devices.

Regardless of the method used to measure performance, the operating parameters must be set for the anticipated operating environment. Key parameters include:

- Input signal level
- System/display line rate
- Luminance (brightness)
- Contrast
- Image size
- Aspect ratio

If the display is used in more than one environmental condition-such as under day and night conditions-a set of measurements is appropriate for each application.

Viewing Environment Considerations

The environment in which a video display device is viewed is an important criterion for critical viewing situations. Applications in which color purity and adherence to set standards are important require a standardized (or at least consistent) viewing environment.

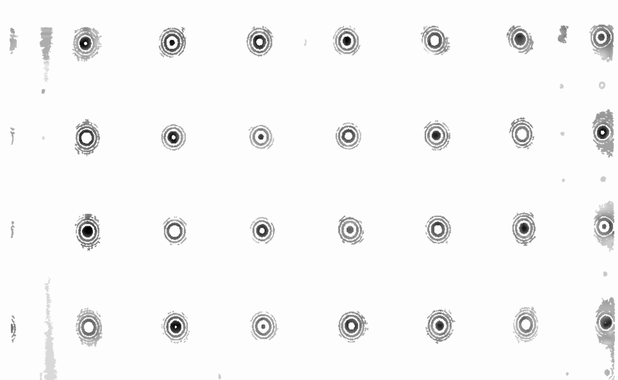


Figure 7.2-29. A 2-axis frequency sweep in which the range of frequencies is swept several times in each axis. Complex patterns such as this may be created for specific test requirements. *(Courtesy of Tektronix.)*

For example, textile colors viewed on a display with a white surround will appear different than the same colors viewed with a black surround. By the same token, different types of ambient lighting will make identical colors appear different on a given display.

SMPTE has addressed this issue with recommended practice (RP) 166-1995, which specifies the environmental and surround conditions that are required in television or video program review areas for the "consistent and critical evaluation" of conventional television signals.[5] Additionally, the practice is designed to provide for repeatable color grading or correction. A number of important parameters are specified in RP 166-1995, including the following:

- The distance of the observer from the monitor screen should be 4 to 6 picture heights for SDTV displays
- The observer should view the monitor screen at a preferred angle in both the horizontal and vertical planes of 0° ± 5° and, in any event, no greater than ±15° from the perpendicular to the midpoint of the screen
- The viewing area decor should have a generally neutral matte impression, without dominant colors
- Surface reflectances should be nonspecular and should not exceed 10% of the peak luminance value of the monitor white.

The *Recommended Practice* suggests placing the monitor in a freestanding environment 2.5-5 screen heights in front of the wall providing the visual surround. Another acceptable approach is to mount the monitor in a wall with its face approximately flush with the surface of the wall. It is further recommended that all light sources in use during picture assessment or adjustment have a color quality closely matching the monitor screen at reference white (D65).

It is often necessary to have black-and-white monitors surrounding one or more color monitors in a studio control room. According to RP 166-1995, the black-and-white monitors should be the same color temperature as the properly adjusted color monitors, 6500 K. Black-and-white monitors are normally equipped with P4 phosphors, at about 9300 K. This cooler color temperature prevents the background surrounding the color monitors from remaining neutral. Most black-and-white monitors can be ordered with 6500 K phosphors.

Picture Monitor Alignment

The proper adjustment and alignment of studio picture monitors is basic to video quality control. Uniform alignment throughout the production chain also ensures consistency in color adjustment, which facilitates the matching of different scenes within a program that may be processed at different times and in different facilities. SMPTE has addressed this requirement for conventional video through RP 167-1995. The *Recommended Practice* offers a step-by-step process by which color monitors can be set. Key elements of RP 167-1995 include the following.[6]

- **Initial conditions.** Setup includes allowing the monitor to warm up and stabilize for 20 to 30 minutes. The room ambient lighting should be the same as it is when the monitor is in normal service, and several minutes must be allowed for visual adaptation to the operating environment.
- **Initial screen adjustments.** The monitor is switched to the setup position, in which the red, green and blue screen controls are adjusted individually so that the signals are barely visible.
- **Purity.** Purity, the ability of the gun to excite only its designated phosphor, is checked by applying a low-level flat-field signal and activating only one of the three guns at a time. The display should have no noticeable discolorations across the face.
- **Scan size.** The color picture monitor application establishes whether the overscan or underscan presentation of the display will be selected. An underscanned display is one in which the active video (picture) area, including the corners of the raster, is visible within the screen mask. Normal scan brings the edges of the picture tangent to the mask position. Overscan should be no more than 5%.
- **Geometry and aspect ratio.** Display geometry and aspect ratio are adjusted with the crosshatch signal by scanning the display device with the green beam only. Correct geometry and linearity are obtained by adjusting the pincushion and scan-linearity controls so that the picture appears without evident distortions from the normal viewing distance.
- **Focus.** An ideal focus target is available from some test signal generators; if it is unavailable, multiburst, crosshatch or white noise can be used as tools to optimize the focus of the displayed picture.
- **Convergence.** Convergence is adjusted with a crosshatch signal; it should be optimized for either normal scan or underscan, depending upon the application.
- **Aperture correction.** If aperture correction is used, the amount of correction can be estimated visually by ensuring that the 2T pulse has the same brightness as the luminance bar or the multiburst signal when the 3 and 4.2 MHz bursts have the same sharpness and contrast.
- **Chrominance amplitude and phase.** The chrominance amplitude and phase are adjusted using the SMPTE color bar test signal and viewing only the blue channel. Switching off the comb filter, if it is present, provides a clear blue channel display. Periodically, the red and green channels should be checked individually in a similar manner to verify that the decoders are working properly.[9]
- **Brightness, color temperature and gray scale tracking.** The 100-IRE window signal is used to supply the reference white. Because of typical luminance shading limitations, a centrally placed PLUGE signal is recommended for setting the monitor brightness control.[10] The black set signal provided in the SMPTE color bars also can be used for this purpose.
- **Monitor matching.** When color matching two or more color monitors, the same alignment steps should be performed on each monitor in turn. Remember, however, that monitors cannot be matched without the same phosphor sets, similar display uniformity characteristics and similar sharpness. The most noticeable deviations on color monitors are the lack of uniform color presentations and brightness shading. Color matching of monitors for these parameters can be most easily assessed by observing flat-field uniformity of the picture at low, medium and high amplitudes.[6]

As more experience is gained with DTV based systems, operating parameters such as those detailed in this section will no doubt be updated to take into consideration the unique attributes and requirements of HDTV.

Picture-Quality Measurements for Digital Television

Picture quality measurement methods include subjective testing, that is always used-at least in an informal manner-and objective testing, which is most suitable for system performance specification and evaluation.[7] A number of types of objective measurement methods are possible for digital television pictures, but those using a human visual system model are the most powerful.

As illustrated in Figure 7.2-30, three key testing layers can be defined for the modern television system:

- **Video quality.** This consists of signal quality and picture quality.
- **Protocol analysis.** Protocol testing is required because the data formatting can be quite complex and is relatively independent of the nature of the uncompressed signals or the eventual conversion to interfacility transmission formats. Protocol test equipment can be both a source of signals and an analyzer that locates errors with respect to a defined standard and determines the value of various operational parameters for the stream of data.
- **Transmission system analysis.** To send the video data to a remote location, one of many possible digital data transmission methods may be used, each of which imposes its own analysis issues.

Table 7.2-1 lists several dimensions of video quality measurement methods. Key definitions include the following:

- **Subjective measurements.** The result of human observers providing their opinions of the video quality.
- **Objective measurements.** Performed with the aid of instrumentation, manually with humans reading a calibrated scale or automatically using a mathematical algorithm.
- **Direct measurements.** Performed on the material of interest, in this case, pictures (also known as picture-quality measurements).
- **Indirect measurements.** Made by processing specially designed test signals in the same manner as the pictures (also known as signal quality measurements). Subjective measurements are performed only

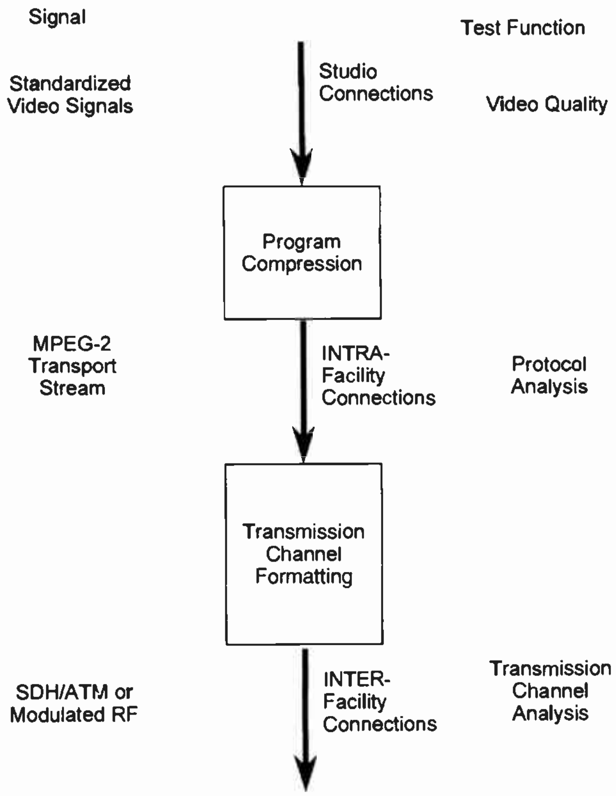


Figure 7.2-30. Video testing layers for digital television.[7]

in a direct manner because the human opinion of test signal picture quality is not particularly meaningful.
- **In-service measurements.** Made while the program is being displayed, directly by evaluating the program material or indirectly by including test signals with the program material.
- **Out-of-service.** Appropriate test scenes are used for direct measurements and full-field test signals are used for indirect measurements.

In the mixed environment of compressed and uncompressed signals, video quality measurements consist of two parts: signal quality and picture quality.

Signal/Picture Quality

Signal quality measurements are made with a suite of test signals as short as one line in the vertical interval.[7] In a completely uncompressed system, such testing will give a good characterization of picture quality. This is not true, however, for a system with compression encoding/decoding because picture quality will change based on the data rate, complexity, and encoding algorithm. Picture quality measurements, instead, require natural scenes (or some equivalent thereof) that are much more complex than traditional test signals. These complex scenes stress the capabilities of the encoder, resulting in nonlinear distortions that are a function of the picture content.

Out-of-service picture quality measurements are similar to indirect signal quality measurements in one aspect: the determination of the system response to a high quality reference. However, they actually measure the degradation in reference picture quality rather than that of a synthetic test signal. In-service, such signals determine the response to program material and its degradation through the system. If the program material is easy, the measurement may not have a great deal of practical value.

Objective indirect signal quality measurements are a reasonably good way to determine the picture quality for uncompressed systems. That is, there is a good correlation between subjective measurements made on pictures from the system and objective measurements made on a suite of test signals using the same system (see Figure 7.2-31). The correlation is not perfect for all tests, however. There are distortions in composite systems, such as false color signals caused by poorly filtered high frequency luminance information being detected as chroma. These distortions are not easily measured by objective means. Also, there are objective measurements that are so sensitive they do not directly relate to subjective results. However, such objective results often are useful because their effect will be seen by a human observer if the pictures are processed in the same way a number of times.

The use of digital compression has expanded the types of distortions that can occur in the modern television system. Because signal quality measurements will not do the job, objective picture quality measurements are needed, as illustrated in Figure 7.2-32. The total picture quality measurement space has increased because of subjective measurements that now include

Table 7.2-1
Video Quality Definitions.[7]

Parameter	In-Service	Out-of-Service
Indirect measurement		
Objective signal quality	Vertical interval test signals	Full-field test signals
Direct measurement		
Subjective picture quality	Program material	Test scenes
Objective picture quality	Program material	Test scenes

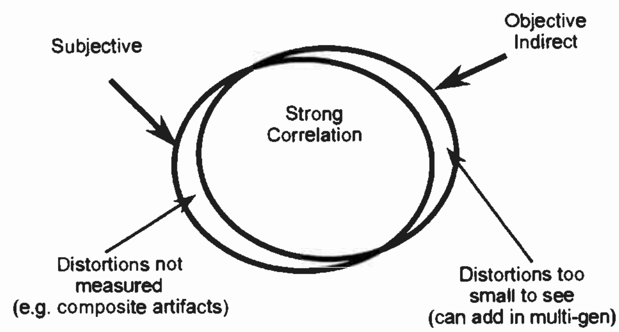


Figure 7.2-31. Functional environment for traditional video measurements.[7]

multi-minute test scenes with varying program material and variable picture quality. The new objective measurement methods must have strong correlation with subjective measurements and cover a broad range of applications.

Even with all the objective testing methods available for analog and full bandwidth digital video, it is important to have human observation of the pictures. Some impairments are not easily measured, yet are obvious to a human observer. This situation certainly has not changed with the addition of modern digital compression. Therefore, casual or informal subjective testing by a reasonably expert viewer remains an important part of system evaluation and/or monitoring.

Automated Picture Quality Measurement

Objective measurements can be made automatically with an instrument that determines picture degradation through the system.[7] Two somewhat mutually exclusive ways are available for classifying objective picture quality measurement systems. Although there are several practical methods with a variety of algorithmic approaches, they may be divided into the following classes:

- **Feature extraction.** An essentially independent analysis of input and output pictures.
- **Picture processing.** The complete input and output pictures are directly compared in some manner and must be available at the measurement instrument.

From a system standpoint, the *glass box* approach utilizes knowledge of the compression system to measure degradation. An example would be looking for blockiness in a DCT system. The *black box* approach makes no assumptions about operation of the system.

In feature extraction, analysis time is reduced by comparing only a limited set of picture characteristics.

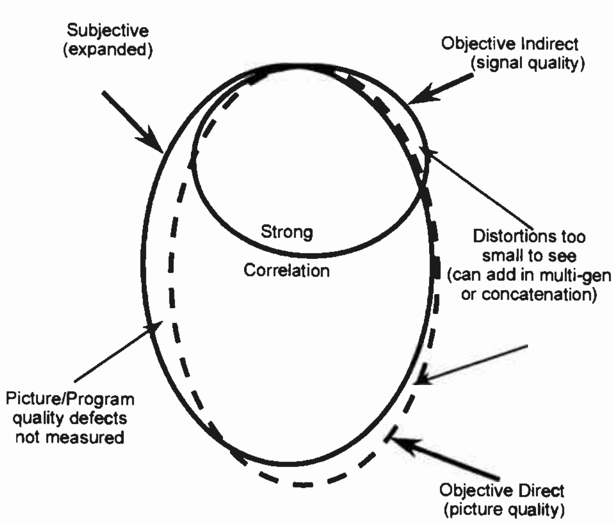


Figure 7.2-32. Functional environment for digital video measurements.[7]

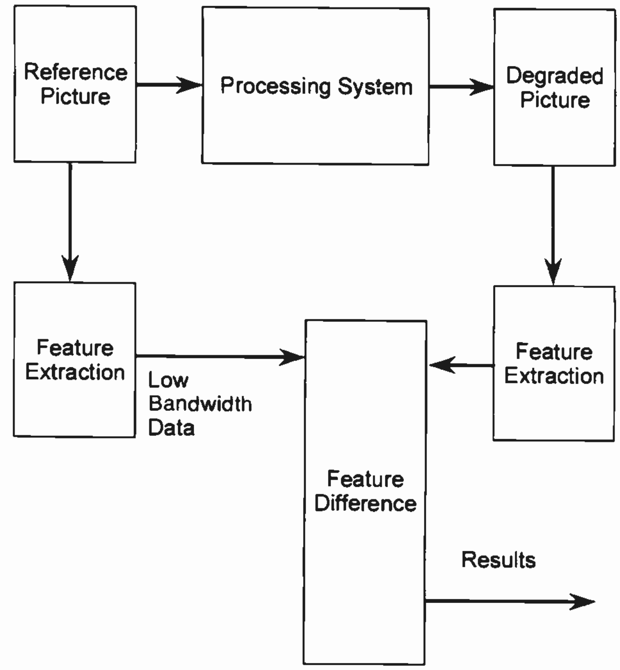


Figure 7.2-33. Block diagram of the feature-extraction method of picture-quality analysis.[11]

These characteristics are calculated and the modest amount of data is embedded into the picture stream. Examples of compression impairments would be block distortion, blurring/smearing or mosquito noise. At the receiver, the same features of the degraded picture are calculated, providing a measure of the differences for each feature (see Figure 7.2-33). The major weakness of this approach is that it does not provide correlation between subjective and objective measurements across a wide variety of compression systems or source material. For the picture processing scheme, the reference and degraded pictures are filtered in an appropriate manner, resulting in a data set that may be as large as the original pictures. The difference between the two data sets is a measure of picture degradation, as illustrated in Figure 7.2-34.

The need for a standard for objective measurements has been addressed in ANSI T1.801.03.[8] (See reference 7 for several variations and improvements to the basic ANSI toolbox.) Because of the limitations of the ANSI method, work continues with the goal of refining objective measurement methods.

A number of approaches have been proposed by researchers using the human visual system (HVS) model as a basis. Such a model provides an image quality metric that is independent of video material, specific types of impairments and the compression system used. The study of the HVS has been going on for decades, investigating such properties as contrast sensitivity, spatio-temporal response and color perception. Perhaps one of the best-known derivatives of this work is the JNDmetrix (Sarnoff/Tektronix).[9]

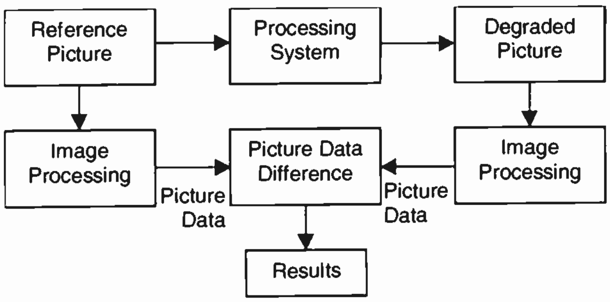


Figure 7.2-34. Block diagram of the picture differencing method of picture-quality analysis.[11]

Serial Digital Bit Stream Analysis

The creation, processing, storage and transmission of video in a digital form has numerous, well-documented advantages over analog signals. It is no surprise, therefore, that use of the serial digital interface (SDI) and serial digital transport interface (SDTI) extension to move signals within and among facilities is increasing every year, to the point now where it is commonplace. As with all good things, however, there are a few hidden problems lurking in the background.

Specifications for the interconnection of digital video signal paths are established with the purpose of setting limits for deviation that will permit proper operation under a variety of conditions. Like analog signal specifications, digital signal specs also establish the limits within which the sending equipment must operate and the receiving equipment must accept.[10] In the digital environment, small violations of the specified limits will usually cause no detrimental affect in the resulting image; this is the nature of digital information transfer, and one of its better known attributes.

The SDI, naturally, concerns itself with transfer of data from one point to another. It follows, then, that three elements must be considered in an analysis of system reliability:

- Transmitter
- Receiver
- Interconnecting medium

If we accept that professional digital video products meet the required SDI specifications (a fair assumption), then the interconnecting medium must be the focus of attention when SDI system design and maintenance is concerned. Given the foregoing, the point of attention is the receiver, or more correctly put, the signal delivered to the input terminals of the receiver.

A digital receiving device has a latitude of acceptable variation within which reliable recovery of data will occur (see Figure 7.2-35). Variations within the high state and low state (as shown) will cause no lost information. Variations greater that the stated specifications may also permit complete information recovery if the performance of the receiver is superior to the nominal (minimum and/or maximum) specification. However, as the state change excursions extend beyond the low state and high state tolerance bands, the performance of the receiver will become unpredictable. Eventually, catastrophic errors will be generated that cannot be masked or recovered by the system.

Reliable operation of a video facility dictates that safety margins be measured and documented, and-in some cases-monitored on a prescribed schedule. The latter case would apply to mission critical links, such as inter-facility lines that would take the plant down if they failed. Also, systems subject to physical stress, such as remote truck equipment, would qualify for regular signal quality analysis.

SMPTE RP 259M

SMPTE RP 259M, the 10 bit 4:2:2 component digital video protocol, produces a signal of 270 Mbps. RP 259M, thus, provides all of the bandwidth necessary to accommodate 8 or 10-bit ITU-R Rec. 601 video, with audio if required, in real time and using conventional 75 Ω coaxial cable.[11] (See reference 12 for the specifications for SMPTE 259M.) Key among the specifications are the peak-to-peak value of 0.8 V and the rise time (transition time) of 0.75 to 1.5 ns. If the signal transmission path had infinite bandwidth and no group delay, the 259M signal would appear as a perfect square wave pulse train. No path is ideal, of course, and herein lies the potential for problems.

As stated previously, the weak link in an SDI system is the path from the transmitter to the receiver, typically several runs of coax interconnected through a router and/or patch bay. The cable itself represents the greatest problem potential when long path lengths are required. Coax can be modeled as an infinite network of inductive and resistive components in series, with distributed shunt capacitance. This, in effect, describes a low-pass filter whose poles increase in number and whose corner frequency moves closer to zero with extending length. Such attenuation with increasing frequency and distance can deteriorate the SDI signal to the point that it becomes unusable.

It is intuitive that different video signals result in different data patterns in a digital system. Consider the case of successive "ones," as illustrated in Figure 7.2-36. This condition results in a true square wave of 50% duty cycle (due to the encoding specified by

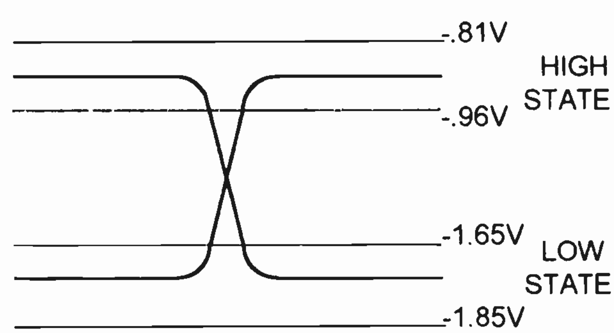


Figure7.2-35. Example of valid data acceptance bands for a digital signal.[10]

RP 259M). Real-life video, of course, results in a unpredictable variety of ones and zeros, producing-in effect-rectangular square waves whose duty cycle is less than 50%.[13] This condition requires a more dense spectrum of harmonics to property define. It follows that considerable low and high frequency harmonics will be present in the SDI signal (significant energy can extend to beyond 1 GHz).

Jitter

Jitter is a related distortion mechanism that may be observed in the SDI/SDTI signal. Jitter is defined as the difference in timing between where a data transition should occur and where it actually does occur. As illustrated in Figure 7.2-37, imperfections in the generation and transmission of the data stream can result in displacement of the transition points to either before or after their proper locations. This timing offset can remain relatively stable, or oscillate between two or more points in time. The latter case is what most engineers consider to be jitter. Minimizing jitter is critical to the performance of SDI based systems.

In order for the receiver to be able to decode the logic levels of the SDI signal, a clock is recovered from the data stream. This recovered clock is used to facilitate decoding of the received data. Because the data transfer is asynchronous, sufficiently large variations in time can result in received data being incorrectly interpreted. Lost data is the result.

The Serial Digital Cliff

Generally speaking, the recovered signal from an SDI/SDTI link is either perfect or basically worthless. An SDI link that is experiencing zero errors, or just a few errors, is considered to be on the operational plateau of the SDI reliability curve. As the link is extended or the S/N otherwise degrades, the system moves forward to the error cliff. As the link progresses over the knee of this cliff, errors climb rapidly to a point sufficient to swamp error control mechanisms built into the SDI system. The path, thus, becomes unusable. Avoiding this well-documented cliff effect requires careful attention to system planning, installation, maintenance and on-going quality control.

Measurement of the S/N of the principle spectral elements in the SDI bit stream is an effective way to

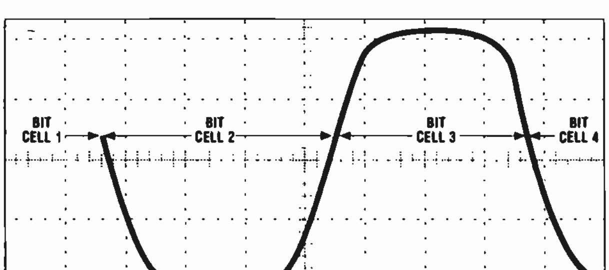


Figure 7.2-36. A portion of a SMPTE 259M datastream showing three successive ones.[13]

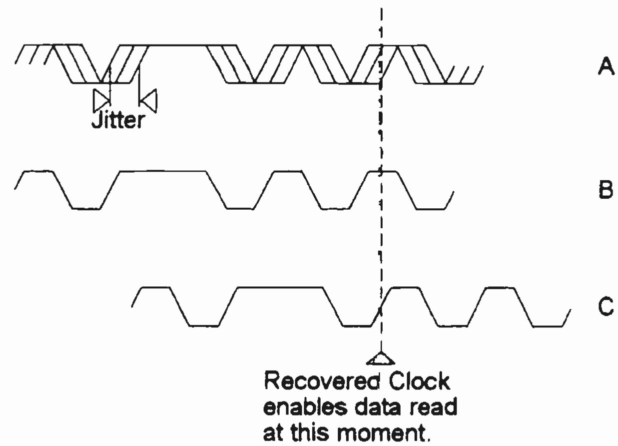


Figure 7.2-37. Representation of jitter: (a) jitter present on the measured waveform, (b) data correctly centered on the recovered clock, (c) signal instantaneously advanced as a result of jitter.[10]

determine how far a path is from the error cliff. For ITU-R Rec. 601 SDI, experience has shown that there are two spectral components whose S/N values are useful in determining the overall health of the link.[17] These are the fundamental frequency of 135 MHz and its third harmonic (405 MHz). It should be noted that because of the coding method used, the fundamental frequency for a component SDI link is not 270 Mbps, but one-half of that value (135 Mbps).[13]

The third harmonic is easy to observe with a spectrum analyzer. At the output of most SDI drivers, the third harmonic starts approximately 35 dB above the noise floor. This compares with the fundamental 135 MHz frequency at 45 to 50 dB above noise. Tests demonstrate that after this signal has passed through approximately 300 meters (1,000 ft) of high quality coaxial cable, the third harmonic is typically down to 8 to 10 dB above the noise floor. As the third harmonic signal approaches 6 dB above the noise floor, clock recovery becomes unreliable and large numbers of errors begin to occur.

As the error cliff is approached, the displayed error rate at the receiver may increase from one per day to one per frame over a S/N range difference of just 3 dB or less. Such a link can become unusable even though the level of the fundamental experiences only modest attenuation through the path.

Pathological Testing

There are a number of tools that can be useful in determining how close a path is to the knee of the cliff. One readily available method is the use of pathological test signals. As outlined previously, the receiver circuitry of an SDI/SDTI link must regenerate the clock signal. To facilitate decoding, most SDI receivers incorporate a signal equalizing circuit to boost the high frequencies of the incoming waveform. This permits easier clock regeneration and data-value determination. Pathological test signals produce bitstreams that stress these circuits.

There are a large number of signal forms that fall under the general category of pathological testing. One common signal stresses the clock regeneration and equalizing circuits by producing values for C and Y that force the SDI bit scrambling circuits to produce a run of 19 zeros and a single one approximately every frame. Another common signal puts the values of C and Y such that a run of 20 ones, followed by 20 zeros periodically, is produced. There are-in fact-thousands of possible C and Y combinations that will stress the receiver.

Measurements under real-world conditions have found that a path will fail under testing with a pathological signal at received levels approximately 2 dB higher than where a typical program (non-pathological) signal will fail.[13] Such tests, therefore, can help to identify whether a given SDI path is at or near the error cliff.

Measuring Jitter

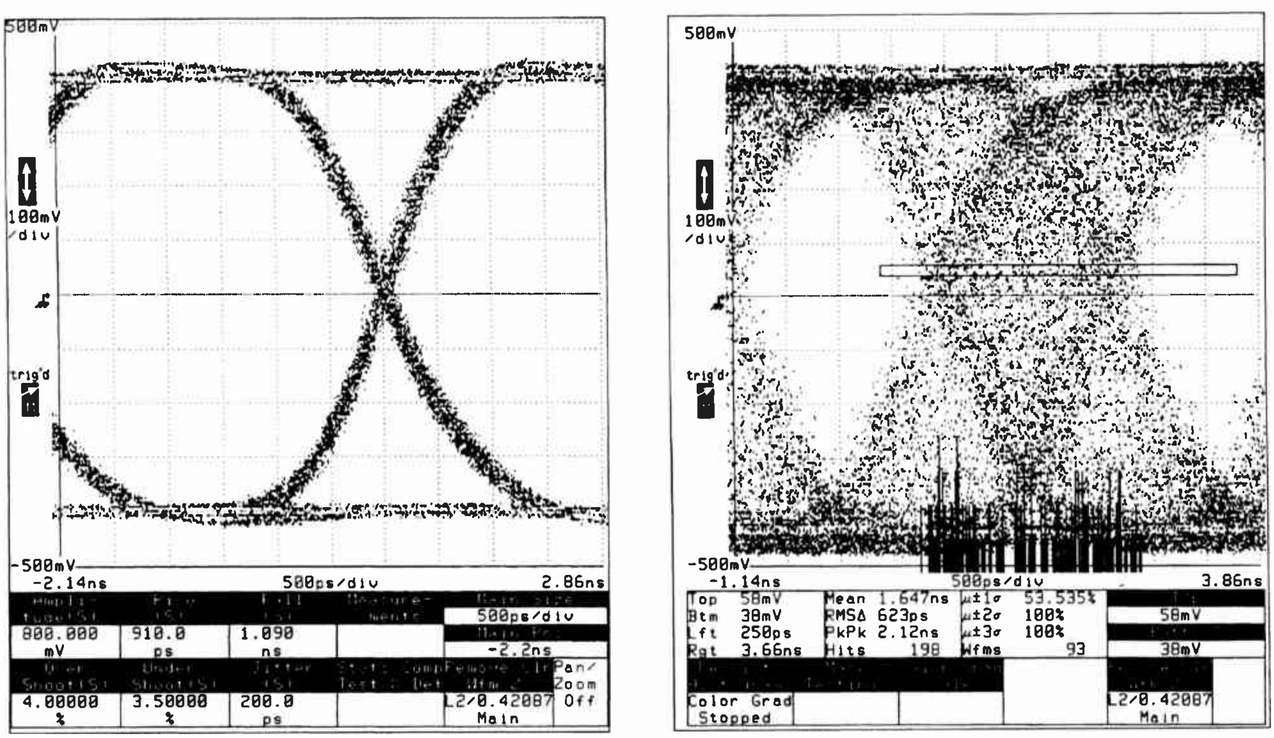
Test equipment is available to measure and characterize the amount of jitter in a link. Several approaches can be taken. One, recommended in SMPTE RP-184, extracts a clock signal from the data stream, which is divided by a given value and used to trigger an eye pattern display. The divisor is typically the same value as the word size (10 bits). This method will, thus, mask any word-related jitter, which is usually quite small. Figure 7.2-38 illustrates jitter measurement using a digitizing oscilloscope.

Quantifying Errors

Errors in digital systems are usually quantified by the bit error rate (BER), which is simply the ratio of bits in error to total bits. Table 7.2-2 gives the BER for one error over different lengths of time for various television systems.[14] BER is a useful measure of system performance where the S/N at the receiver is such that noise-produced random errors occur.

Bit scrambling is used in the SDI system to lower the dc content of the signal and to provide sufficient zero crossings for reliable clock recovery at the receiver.[12] It is the nature of the descrambler that a single bit error will cause an error in two words (samples). Furthermore, there is a 50% probability that the error will occur in one of the words being in the most significant, or next to the most significant, bit position. The resulting error rate of 1 error/frame will be noticeable by a reasonably patient observer. This situation, clearly, is unacceptable in professional video applications.

Figure 7.2-39 shows the block diagram of a basic serial digital transmitter and receiver system. The intuitive method of testing the serial link is to add cable to the point where the link is unusable. Because coax is not itself a significant source of noise, it is the noise figure (NF) of the receiver that will determine the basic operating S/N of the system. Assuming an automatic equalizer in the receiver (which is usually the case), as more cable is added, eventually the signal level resulting from coax attenuation will cause the



(a) (b)

Figure 7.2-38. Measurement of jitter using a digitizing oscilloscope: (a) low jitter, absolute jitter = 200 picoseconds p-p; (b) significant jitter present, absolute jitter = 2.12 ns p-p. *(Courtesy of Tektronix.)*

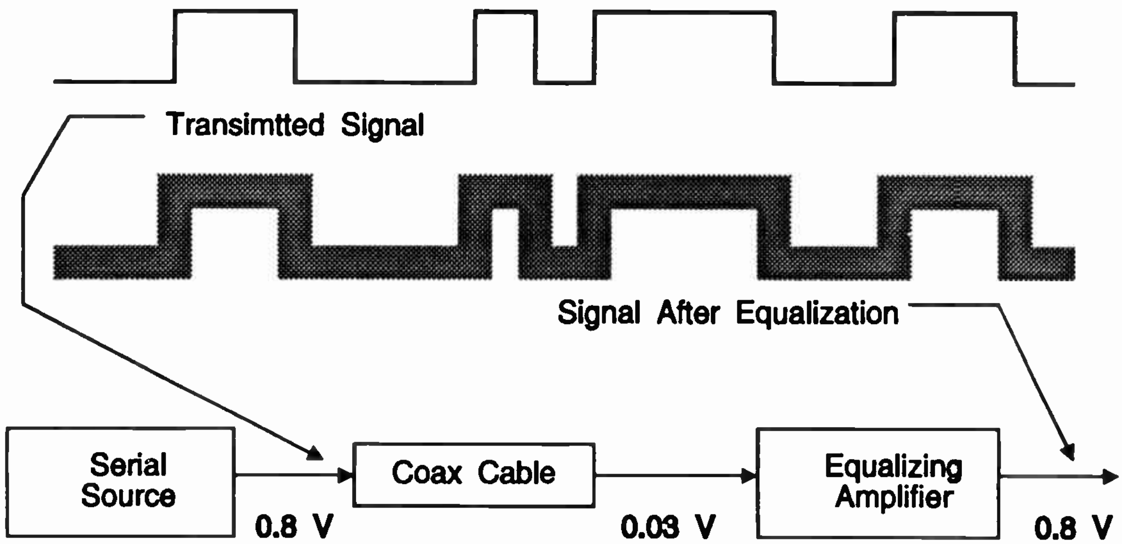


Figure 7.2-39. Basic serial digital transmitter/receiver system. Note typical voltage levels (for composite video) and the function of the equalizing amplifier.[14]

S/N in the receiver to degrade to the point that errors occur.

Based on the scrambled non-return-to-zero-inverted (NRZI) channel code used with SDI and assuming gaussian-distributed noise, a calculation using the error-function provides theoretical values for error rate as a function of S/N, as shown in Table 7.2-3.[14] An examination of the table will show that for composite (NTSC) serial digital transmission, a 4.7 dB increase in S/N changes the resulting condition from 1 error/frame to 1 error/century. For composite digital, the calibration point for the calculation is 400 m of Belden 8281 coax. Other types of cable can be used with adjustments made for the calculation point, if necessary.

The same theoretical data can be expressed to show error rates as a function of cable length, as given in Table 7.2-4.[14] (This data is reproduced in graph form in Figure 7.2-40). There is a sharp knee in the graph as cable length is extended beyond a certain critical point (380 m, or approximately 1,250 ft, for composite digital video). Similar results are obtained for other standards; the critical point calculation is 360 m for PAL and 290 m for ITU-R Rec. 601 video. Cable lengths and headroom scale proportionally. Good engineering practice would suggest a minimum of 6–8 dB margin for reliable SDI transmission.

Practical systems include equipment that does not necessarily completely reconstitute the signal in terms of S/N. For example, sending the SDI signal through a distribution amplifier or routing switcher may result

Table 7.2-3

Error Rate as a Function of S/N for Composite Serial Digital.[14]

Time Between Errors	BER	SNR (dB)	S/N (volts ratio)
1 microsecond	7×10^{-3}	10.8	12
1 millisecond	7×10^{-6}	15.8	38
1 television frame	2×10^{-7}	17.1	51
1 second	7×10^{-9}	18.1	64
1 minute	1×10^{-10}	19.0	80
1 day	8×10^{-14}	20.4	109
1 month	3×10^{-15}	20.9	122
1 century	2×10^{-18}	21.8	150

Table 7.2-2

Error Frequency and Bit Error Rates.[15]

Time Between	NTSC, 143 Mbits/s	PAL, 177 Mbits/s	Component, 270 Mbps
1 television frame	2×10^{-7}	2×10^{-7}	1×10^{-7}
1 second	7×10^{-9}	6×10^{-9}	4×10^{-9}
1 minute	1×10^{-10}	9×10^{-11}	6×10^{-11}
1 hour	2×10^{-12}	2×10^{-12}	1×10^{-12}
1 day	8×10^{-14}	7×10^{-14}	4×10^{-14}
1 week	1×10^{-14}	9×10^{-15}	6×10^{-15}
1 month	3×10^{-15}	2×10^{-15}	1×10^{-15}
1 year	2×10^{-16}	2×10^{-16}	1×10^{-16}
1 decade	2×10^{-17}	2×10^{-17}	1×10^{-17}
1 century	2×10^{-18}	2×10^{-18}	1×10^{-18}

Table 7.2-4

Error Rate as a Function of Cable Length Using 8281 Coax for Composite Serial Digital.[14]

Time Between Errors	BER	Cable Length (meters)	Attenuation (dB) at 1/2 Clock Frequency
1 microsecond	7×10^{-3}	484	36.3
1 millisecond	7×10^{-6}	418	31.3
1 television frame	2×10^{-7}	400	30.0
1 second	7×10^{-9}	387	29.0
1 minute	1×10^{-10}	374	28.1
1 day	8×10^{-14}	356	26.7
1 month	3×10^{-15}	350	26.2
1 century	2×10^{-18}	338	25.3

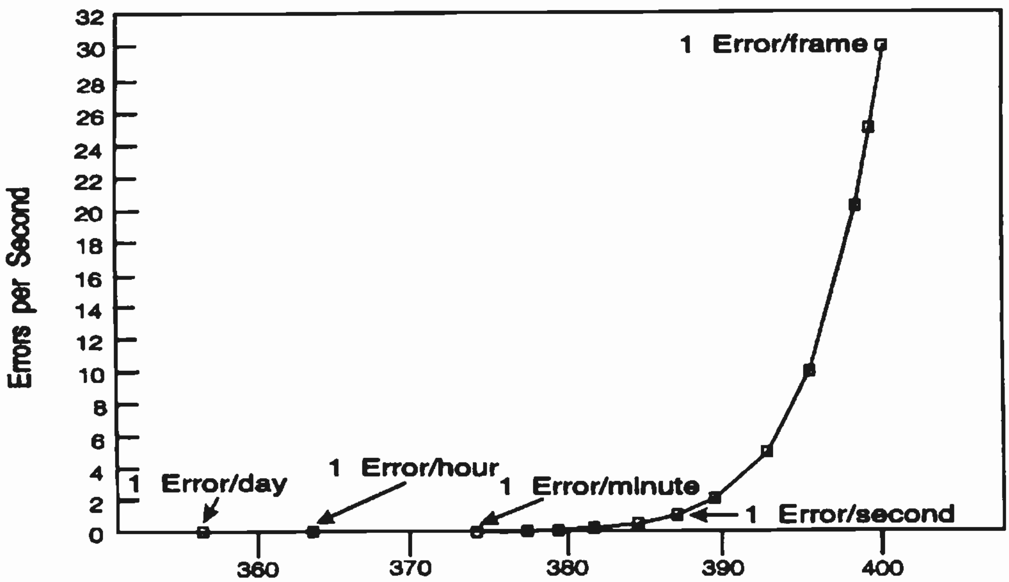


Figure 7.2-40. Error rate as a function of cable length (in meters) for composite video signals on an SDI link.[14]

in a completely useful, but not completely standard signal that is sent to a receiving device. The characteristics of this non-standard signal could include both jitter and noise. The sharp knee characteristic of the system, however, would remain, occurring at a different amount of signal attenuation.

Eye Diagram

Most engineers are familiar with an oscilloscope display of repetitive waveforms such as sine, square or triangle waves. These are known as single value displays because each point on the time axis has only a single voltage value associated with it.[16]

When analyzing a digital telecommunications waveform, single value displays are not very useful. Real communications signals are not repetitive, but instead consist of random or pseudorandom patterns. A single value display can only show a few of the many different possible one/zero combinations. A number of pattern dependent problems will be overlooked if they do not occur in the small segment of the waveform appearing on the display of the conventional oscilloscope.

The eye diagram overcomes the limitations of the single value display by overlapping all of the possible one/zero combinations on the screen. Eye diagrams are multivalued displays in that each point on the time axis has multiple voltage values associated with it.

The eye diagram has become a valuable tool in analyzing digital communications systems because of the unique properties that it is capable of displaying.

REFERENCES

1. Tannas, Lawrence E., Jr., Flat Panel Displays and CRTs, Van Nostrand Reinhold, New York, pg. 18, 1985.
2. Standards and Definitions Committee, Society for Information Display.
3. "Broadening the Applications of Zone Plate Generators," Application Note 20W7056, Tektronix, Beaverton, OR, 1992.
4. Verona, Robert, "Comparison of CRT Display Measurement Techniques," Helmet-Mounted Displays III, Thomas M. Lippert, (ed.), Proc. SPIE 1695, SPIE, Bellingham, WA, pp. 117–127, 1992.
5. "Critical Viewing Conditions for Evaluation of Color Television Pictures," SMPTE Recommended Practice RP 166-1995, SMPTE, White Plains, NY, 1995.
6. "Alignment of NTSC Color Picture Monitors," SMPTE Recommended Practice RP 167-1995, SMPTE, White Plains, NY, 1995.
7. Fibush, David K., "Picture Quality Measurements for Digital Television," Proceedings of the Digital Television '97 Summit, Intertec Publishing, Overland Park, KA, December 1997.
8. ANSI Standard T1.801.03-1996, "Digital Transport of One-Way Video Signals: Parameters for Objective Performance Assessment," ANSI, Washington, DC, 1996.
9. Fibush, David K., "Practical Application of Objective Picture Quality Measurements," Proceedings IBC '97, IEE, pp. 123–135, Sep. 16, 1997.
10. Finck, Konrad, "Digital Video Signal Analysis for Real-World Problems," in NAB 1994 Broadcast Engineering Conference Proceedings, National Association of Broadcasters, Washington, DC, pg. 257, 1994.
11. Haines, Steve, "Serial Digital: The Networking Solution?," in NAB 1994 Broadcast Engineering Conference Proceedings, National Association of Broadcasters, Washington, DC, pg. 270, 1994.

12. Standard SMPTE 259M, "Serial Digital Interface for 10-bit 4:2:2 Components and 4Fsc NTSC Composite Digital Signals."
13. Boston, J., and J. Kraenzel, "SDI Headroom and the Digital Cliff," Broadcast Engineering, Intertec Publishing, Overland Park, KA, pg. 80, February 1997.
14. Fibush, David K., "Error Detection in Serial Digital Systems," NAB Broadcast Engineering Conference Proceedings, National Association of Broadcasters, Washington, DC, pp. 346–354, 1993.
15. Stremler, Ferrel G., "Introduction to Communications Systems," Addison-Wesley Series in Electrical Engineering, Addison-Wesley, NY, December 1982.
16. "Eye Diagrams and Sampling Oscilloscopes," Hewlett-Packard Journal, Hewlett-Packard, Palo Alto, CA, pp. 8–9, December 1996.

7.3
RADIO FREQUENCY SIGNAL ANALYSIS

DONALD MARKLEY
MARKLEY & ASSOCIATES, PEORIA, IL

INTRODUCTION

The analysis of radio frequency (RF) signals has benefited greatly from the newer generation of digital test equipment. In particular, the digital spectrum analyzer and vector network analyzer have changed the way in which transmitted signals and RF networks can be analyzed and adjusted.

In its simplest form, the spectrum analyzer can be visualized as a self-tuning receiver with adjustable bandwidth. The operator selects the center frequency around which measurements are desired. The receiver bandwidth and the width of the band of frequencies to be measured are then set. The spectrum analyzer then scans across those frequencies and displays all received signals. With analog systems, the display was sometimes difficult to interpret as certain combinations of sweep characteristics resulted in a display that was not constantly visible. Experienced operators learned to cope with the display and, through experience, could evaluate even the weakest of signals. However, digital spectrum analyzers have made it much easier for the occasional or new user to benefit from the device.

The digital spectrum analyzer functions in a similar fashion to a traditional analog device as far as general measurement principles are concerned. In fact, it is often possible to switch to a totally analog display. However, in the digital display mode, the unit will display each measured signal without fading between signal traces. The display is simply updated each time the system measures the desired frequencies. For transient measurements, the system holds the display as long as desired to show a single trace. Most modern units will also store one or more sets of measurements for plotting or printing using a computer interface.

An additional feature of the digital spectrum analyzer is the ability to store the maximum value of the signal noted on any frequency within the selected band or frequencies over a period of time. This feature is particularly useful in the analysis of spurious signals or the observation of a modulation envelope.

NATIONAL RADIO SYSTEMS COMMITTEE (AM RADIO)

Standard broadcast stations are required to perform an annual measurement of spurious radiation and the occupied bandwidth of their signals. Section 73.44 of the Federal Communications Commission (FCC) Rules specifically identifies the instrument to be used as a suitable swept-frequency RF spectrum analyzer. The instrument must use a 300 Hz resolution bandwidth and perform the measurement over a 10 minute period. The peak signal values at each frequency during that period are to be displayed.

Other specialized receivers may be used but, in the case of any disagreement over the results, the peak storing spectrum analyzer is to be used to determine the correct value. The Rules also specify the manner in which the measurements are to be performed. In particular, the station must be operated normally and the measurements taken at approximately 1 km from the center of the antenna system. The Rules further explain how this is to be done for directional systems.

It is often difficult to obtain enough signal to perform the measurement at the specified 1 km distance, especially for lower powered stations. Loop antennas are available that provide good response and gain characteristics. In addition, the use of a loop antenna can eliminate the signals of other stations through proper orientation. Still, even the loop does not always yield sufficient signal to allow for the dynamic range necessary to show compliance with the limits in the Rules. In such cases, a preamplifier may be added to the system as long as its response does not significantly alter the overall frequency response of the system.

Figure 7.3-1 illustrates the results of an AM modulation envelope measurement. The limits specified by the FCC are shown on the plot. In addition to the basic envelope, the Rules require measurement of the harmonics of the station's assigned frequency. All harmonics are to be either 43 + Log (power in watts) or 80 dB below the carrier, whichever is the lesser attenuation. Stations operating at less then 158 W are only required to show harmonic attenuation of at least 65 dB.

While measurement of the station's harmonics can be performed with the spectrum analyzer, the results are often masked by noise. More accurate results can often be obtained by using a field-strength meter in combination with a calibrated antenna. One significant source of error in measuring harmonics with a spectrum analyzer is the gain of the antenna used. The gain will vary considerably over the band of frequencies from the carrier to the third harmonic. Accurate measurements can only be obtained when the entire system is calibrated. Antennas of known gain are available, but it is much easier to simply use a field-strength meter for such measurements.

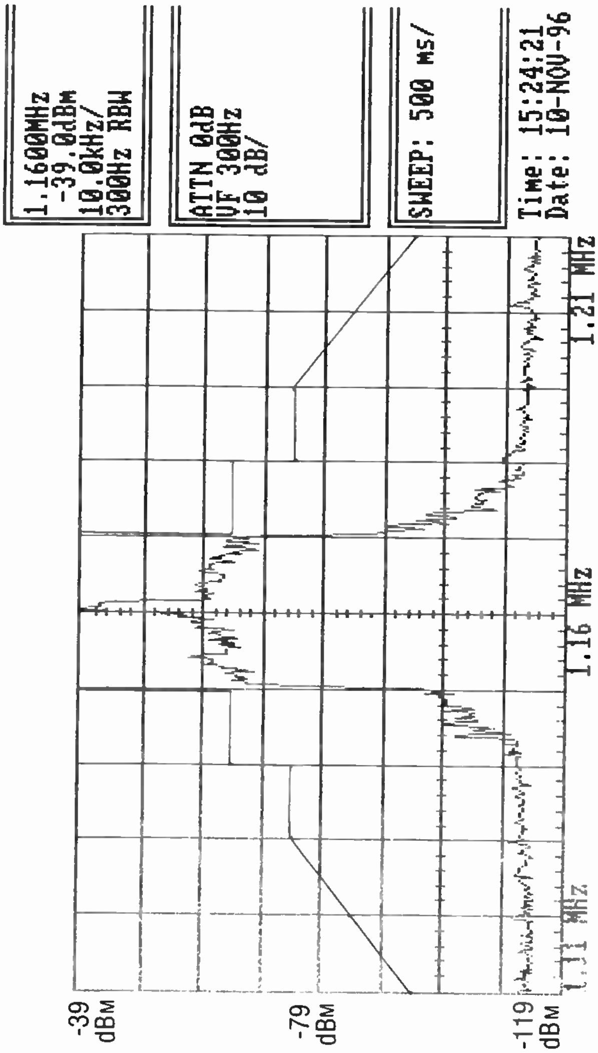


Figure 7.3-1. NRSC measurement of AM transmitter with FCC limits.

TELEVISION SIDEBAND MEASUREMENTS

The FCC has rigorous requirements for the sideband response of television transmitters. The measurements to confirm proper operation, as well as the tuning of the transmitters themselves, require both a spectrum analyzer and a suitable signal generator. The more common method of measurement has involved the use of a sideband adapter which, together with the spectrum analyzer, provides a stable and easily analyzed waveform.

The most common sideband adapter is the Tektronix 1405, although the use of that unit requires a spectrum analyzer that has been modified for such use. The units are interconnected so that their frequencies track. The sideband adapter generates a test signal which sweeps the selected television channel and includes markers to aid in determining the system frequency response. Figure 7.3-2 shows the response of a UHF television

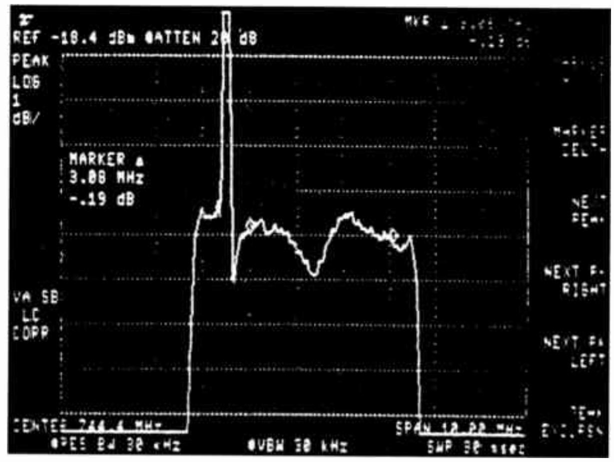


Figure 7.3-2. Television transmitter sideband response.

transmitter using a sideband adapter. In that figure, the vertical scale is 1 dB/div.

An additional method of sweeping the TV transmitter uses a (sin x)/x signal from a digital television test signal generator. The wide frequency response of that signal, together with the accuracy inherent in modern spectrum analyzers, provides sufficient accuracy to meet the FCC's measurement requirements. An additional advantage of this method is that the cost of a separate sideband adapter is avoided.

Tracking Generator Measurements

A general problem exists when a signal generator and spectrum analyzer are used to measure the frequency response of systems. In simplest terms, the signal generator only generates one frequency at any given instant. The generator then sweeps continuously over a desired band of frequencies or, in the case of modern, digitally swept generators, moves in increments across that band. Since the spectrum analyzer can only measure one frequency at a given instant, some means must be found to cause the analyzer to perform measurements on the same frequency that is being generated at that time. This is accomplished with a device known as a *tracking generator*.

The tracking generator is either installed in a spectrum analyzer or mounted externally and connected directly to an analyzer. The two units are synchronized so that the settings on the analyzer controls also determine the method in which the generator operates. The result is that the generator frequency and the frequency measured by the analyzer are the same at all times. For analog equipment, the output of the tracking generator, when connected directly to the input of the spectrum analyzer, will appear as a straight line at the generator output magnitude. For digital equipment, the output of the tracking generator may appear as a row of pulses.

Figure 7.3-3 demonstrates one way this equipment can be used. The device under test is simply connected between the tracking generator output and the input to the spectrum analyzer. Since the magnitude of the tracking generator signal output is known, this test

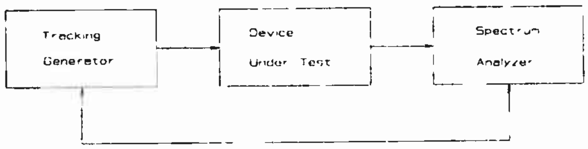


Figure 7.3-3. Block diagram of tracking generator and spectrum analyzer test setup.

setup allows the frequency response of the device to be measured very accurately over the range of interest. It must be noted that this measurement only shows the magnitude of the response. The spectrum analyzer is a scalar device and does not provide vector (magnitude and phase) data.

Using a return loss bridge, the tracking generator and spectrum analyzer combination can measure the return loss from antenna systems. The return loss can in turn be used to calculate voltage standing wave ratio (VSWR) and the reflection coefficient of the load as shown in Equation 1.

$$|\Gamma| = 10^{(RL/20)}$$

$$\text{VSWR} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad [1]$$

Where: Γ = Reflection Coefficient
RL = Return Loss in dB

However, measurements made on a system in this fashion do not provide sufficient information to determine the location of any problems in the system—only the frequency response of the overall system is measured. Determining the location of faults requires either vector measurement information or the use of a time domain reflectometer.

Time Domain Reflectometers

The time domain reflectometer (TDR) applies a pulse or step of voltage to a transmission line or antenna system. An oscilloscope is normally used as the display device and is supplied signals from the TDR. The oscilloscope is usually operated in the *x-y* mode with the horizontal and vertical signal provided directly from the TDR. The oscilloscope displays the measured signal on the transmission line input as a function of time. Some TDRs incorporate an LCD display or a chart recorder.

The applied pulse will travel down the cable at a known rate which is the velocity of propagation of the cable under test (usually available from the cable manufacturer). When the pulse passes an irregularity or discontinuity in the cable, a portion of the pulse will be reflected back to the source. The strength of the reflection will be determined by the impedance change at the point of the discontinuity. Equation 2 can be used to determine the coefficient of reflection at the impedance change. The reflected signal is simply equal to the incident signal, or the pulse, multiplied by the coefficient of reflection.

$$\Gamma = (Z - Z_0)/(Z + Z_0) \qquad [2]$$

Where: Γ = Reflection coefficient
Z = Impedance at discontinuity
Z_0 = Characteristic impedance of line

The reflected signal travels back to the input of the transmission line where the TDR is connected. The distance from the TDR to the irregularity is determined by Equation 3. The factor of ½ in that equation is needed to account for the fact that the signal travels over the distance from the TDR to the fault twice, first in the forward direction and then in the reverse direction. An antenna on the end of the transmission line will normally look like either an open or a short circuit to the TDR pulse—an antenna that has dc continuity will appear to be a short with a reflection coefficient of −1; if the antenna is an open at dc, it will appear to be an open circuit with a reflection coefficient of +1.

$$d = 3 \times 10^2 \times Vp \times t/2 \qquad [3]$$

Where: d = Distance to fault in meters
Vp = Velocity of propagation in cable
t = Time between incident and reflected pulse

A TDR is usually equipped with an adjustable delay calibrated in distance to fault. The initial adjustment of zero distance establishes a reference on the oscilloscope. The delay can then be adjusted to place any reflected signal at the same reference point. The distance to the fault is then read directly from the TDR. The distance to the fault is only accurate if the instrument has been calibrated for the correct velocity of propagation. This is relatively easy to do if the exact length of the cable is known. The velocity of propagation setting can be adjusted until the reflection from the end is at the correct distance. All points between the TDR and the end of the cable will then also be at the correct distance.

Newer TDR equipment may allow for keypad entry of the velocity of propagation value. In some cases, the type of transmission line is selected by the user from the display and the velocity is then automatically selected by the unit. Again, the calibration can be checked easily if the cable length is known.

The sensitivity of TDR measurements is usually limited by the noise level present on the line. With an antenna present at the end of the line, the RF noise level present may mask minor reflections. Two solutions exist to that problem. One is to use a higher powered TDR. The other, easier method is to terminate the transmission line in either an open or a short circuit. The more complete the open or short is made, the smaller the discontinuity that can be seen. A good TDR on a tightly sealed rigid transmission line should be able to see every connection between sections as well as the insulators along the center conductor.

A line that has experienced an arc across one of the insulators will leave a carbon path although that path can be very difficult to find. Obviously, the TDR does not use a sufficiently high voltage to cause the arc to reoccur. If the line can be tightly sealed to lower the noise floor as much as possible, the location of the carbon path may become visible. Faults of this type are very difficult to find when operating at test instrument power levels. Fortunately, their location will probably become quite evident after the normal transmitter power level recreates the arc a few more times.

Situations do exist that will cause highly misleading indications on a TDR. For example, a gradual change in the line impedance (versus distance) will not cause a distinct reflection leading the operator to conclude that no problem exists. This situation can occur when an air dielectric line with a slight slope in the horizontal run accumulates water from being improperly pressurized. The impedance change along the line will be gradual as the level of the water approaches the center conductor. To the TDR, the gradual change does not present a clean point of reflection and the overall change in the line may be missed. To avoid this pitfall, the operator should always try to first identify (using the TDR) the end of the line. If it appears to be missing, a problem obviously exists which the TDR method cannot identify.

NETWORK ANALYZER MEASUREMENTS

The digital network analyzer is essentially an accumulation of pieces of test equipment in one box. This equipment includes a synthesized signal generator, up to three separate receivers, a controller/analyzer system and a display. Depending upon the model, the system may also include a reflection or transmission test set and, in some cases, an "s-parameter" test set. Figure 7.3-4 is a block diagram of a typical network analyzer.

Traditionally, analog network analyzers were swept-frequency devices and were not as capable of compensating for system errors (such as test setup amplitude distortions) as are the newer digital systems. Some attempts were made with the analog instruments to add self-correcting circuits. However, the build-up of system noise, as more and more components were added, tended to reduce the ability of the system to measure very low values of VSWR. The newer digital analyzers have eliminated that problem.

The digital network analyzer does not do swept-frequency measurements. Rather, the band of interest is analyzed on a point-by-point basis. The initial instructions to the system controller include the number of points to be measured. That number of measurement points is then uniformly (or in some cases logarithmically) distributed over the range of interest. For each point, the generator and receivers are set to a selected frequency. Once the generator has locked on frequency, the output is directed to a transmission test set or s-parameter test set. For illustrative purposes, assume that the more common transmission test set is in use. Within the transmission test set, the signal from the generator is divided into two equal signals. One of these is returned to one receiver channel as a reference. The other signal is applied to a return loss bridge or directional coupler. This second signal is then routed to the system under test. The reflected signal then goes to another network analyzer receiver channel for comparison to the reference signal. The difference between those signals represents the reflection coefficient.

In a scalar network analyzer, the system performs the required calculations and the results are displayed as the simple magnitude of the reflection, the return loss in dB or the VSWR. Such a VSWR plot is shown in Figure 7.3-5. Traditionally, the display has been a CRT, although some of the newest models have changed to LCD displays. That change has resulted in a significant reduction in system weight without any reduction in the quality of the presentation. A vector network analyzer goes further in determining both the magnitude and the phase of the reflected signal. That allows a Smith Chart presentation of the measured impedance as shown in Figure 7.3-6. Some systems also permit a display of phase only as well as a polar plot of the system impedance. Markers can be used to identify the critical points such as visual carrier, aural carrier or color carrier frequencies.

A major feature of digital vector network analyzers is the ability to calibrate the unit in the field. To do this, three accurate terminations are required. The normal procedure is to use a short circuit, an open circuit and a calibrated termination. The open and short circuits are placed in their respective connectors at the same location as the load in the resistive termination. In addition, the open is shielded to eliminate any errors due to stray RF signals. For each of the three terminations, the system measures the impedance of the termination at each of the frequencies across the selected range. After the three sets of measurements are completed, a set of correction coefficients is calculated to correct the measured data to accurately represent an open, a short and a perfect 50 Ω resistive load. That correction matrix is used on all subsequent measurements to compensate for system errors. This corrects for cable leakage or resistance problems, errors in the power divider and return loss bridge and other minor system variations. However, the final result is only as accurate as the calibrating devices.

If the network analyzer, after undergoing calibration, determines the system VSWR at some frequency to be 1.05, the result is actually based on the value of the resistive termination being 50 Ω with zero reactance. Any error in the value of the termination causes a corresponding system error in all measurements performed by the system. As a result, the operators of these systems usually guard their terminations zealously, particularly on the road.

The ability to calibrate the system in the field has opened an additional level of measurement ability. In the past, it was necessary to lift the measurement equipment to the top of the tower if measuring the actual antenna was desired (assuming that this discussion refers to a large antenna). With the ability to calibrate the system using the three precision terminations, it is possible to perform the calibration at the

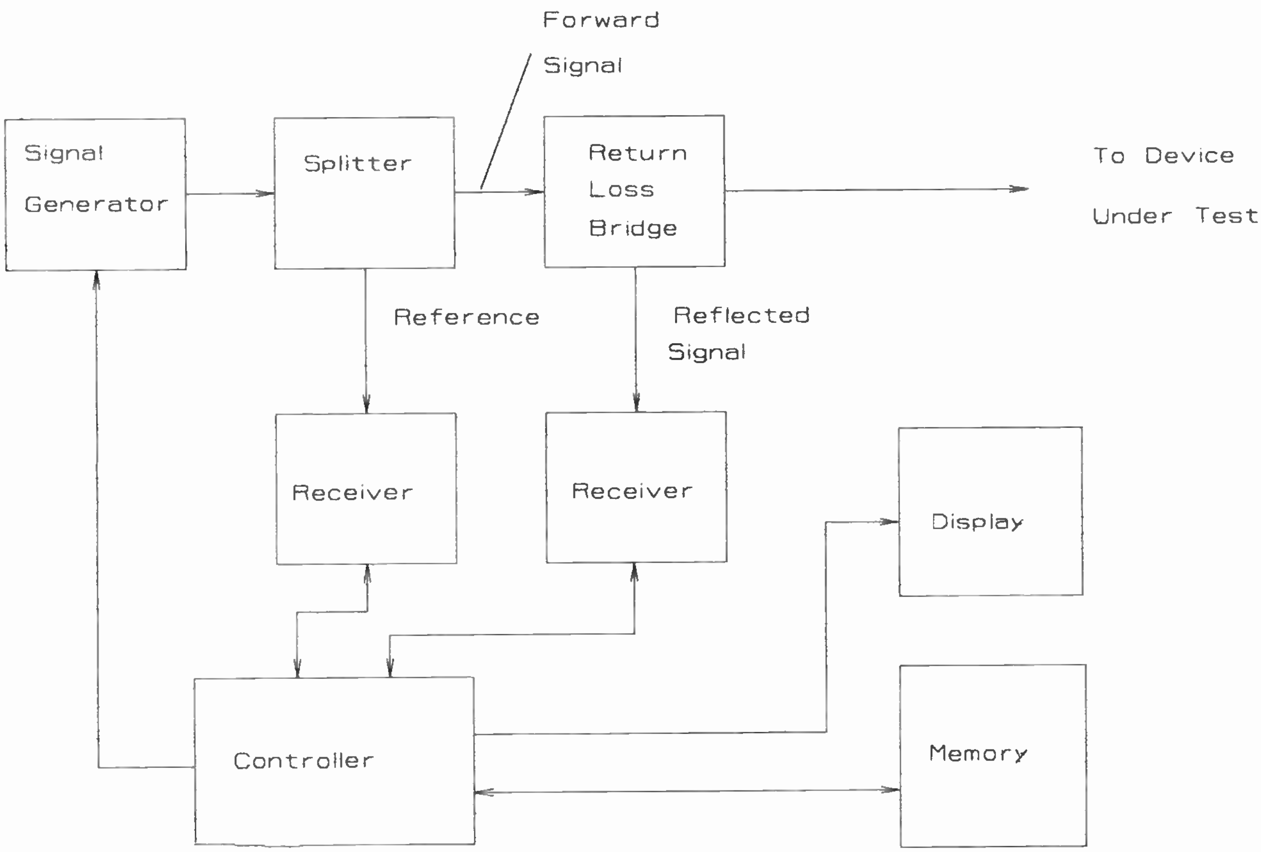


Figure 7.3-4. Generic block diagram of a network analyzer.

load end of the transmission line rather than at the output of the test equipment, correcting for the presence of the transmission line. Once calibrated, the display will now show the actual VSWR of the antenna itself irrespective of the connecting hardware. Due to other developments, to be discussed later, that particular measurement is not as necessary as it was in the past. However, it is still useful when elements of the antenna need to be tuned in the field.

A major problem occurs when performing measurements on antenna systems, particularly in the UHF television band. The adapters used to convert from

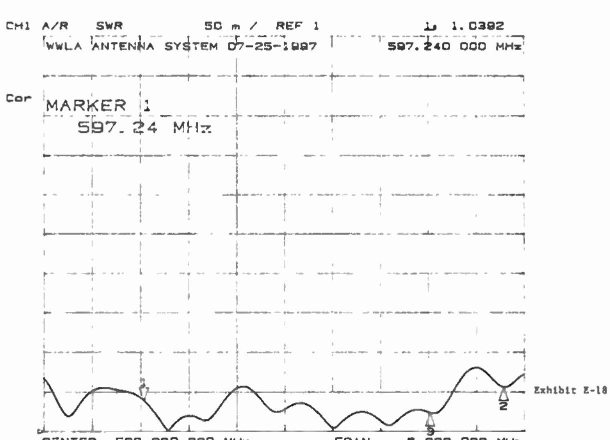


Figure 7.3-5. VSWR plot of TV antenna and transmission line.

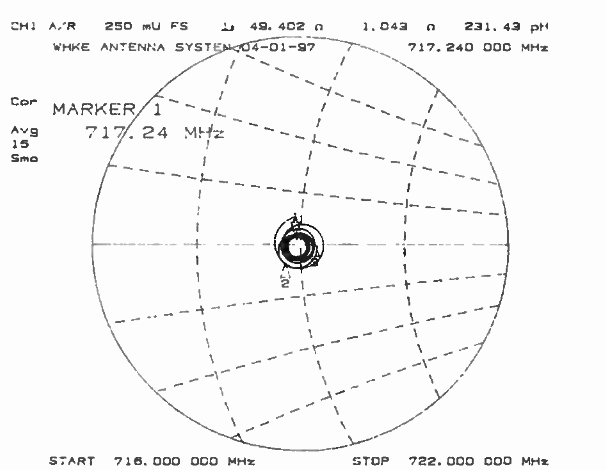


Figure 7.3-6. Smith Chart plot of TV antenna and transmission line with good impedance match at the antenna.

type N connectors to large transmission lines must be carefully tuned to eliminate errors. Without tuning, it is not unusual for adapters (from type N to 6 1/8 in. or larger cable) themselves to introduce a VSWR of 1.1:1 or more. When measuring VSWR system values of less than 1.05:1, such errors in the adapters are unacceptable. When manufactured, these adapters are often tuned using a slotted line, an approach not normally available in the field. In any case, the purpose of all of the initial tuning and calibration is to eliminate the errors caused by the instrumentation itself.

Modern television transmitting antennas are normally tuned for a VSWR of 1.05 or less at visual carrier, 1.08 or less at aural and color carriers and a maximum value of 1.1:1 across the 6 MHz channel. Some antenna manufacturers predict the necessity of tuning DTV antenna systems for a minimum return loss of 30 dB across the channel. That corresponds to a VSWR of 1.065:1. Clearly, good instrument calibration will be important when making measurements of these types.

Simulated TDR Measurements

When tuning transmitting antennas, a TDR does not provide any information concerning the match to the antenna itself. In fact, since the antenna normally appears to be either an open or a short at dc, the antenna will present a very large reflection. On the other hand, measurement of the VSWR with a network analyzer or a spectrum analyzer with a tracking generator will not provide any information concerning the location of a mismatch in the band of frequencies of interest. It is here that another feature of the vector network analyzer becomes of interest. It is possible to determine the distance to faults at RF frequencies and the magnitude of those faults.

When the network analyzer performs the measurement of the reflection coefficient at the selected number of points, it is building a series of values that represent the system response as a function of frequency. This can be said to be a function in the frequency domain. When the Inverse Fourier Transform (Equation 4.) is applied to that measured data, the result is the response expressed in the time domain. That is, the response demonstrates the distance to faults along the line and the magnitude of those faults. For the analysis of TV antenna systems, that response is calculated across the 6 MHz channel and shown as VSWR versus distance.

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(w)c^{iwt}\,dw \qquad [4]$$

where:
$f(t)$ = Function of time
$F(w)$ = Function of frequency
$w = 2\pi f$
f = Frequency in hertz

There are some precautions to be taken in determining the distance to faults by this method. Due to the limitation on the amount of data used in the calculation, some erroneous indications can occur due to a phenomena called aliasing. The maximum distance range to avoid aliasing is determined by Equation 5 and is a function of the frequency span over which the measurements are performed, the number of data points used and the relative propagation velocity of the transmission line.

$$d\,\text{max} = (1.5 \times 10^8)(Np)(Vp)/(F_2 - F_1) \qquad [5]$$

where:
d max = Maximum measurement distance without aliasing
Np = Number of measurement points
Vp = Velocity of propagation
F_1 = Starting frequency
F_2 = Stopping frequency

The most common problem in the analysis of TV transmitting antenna systems is the mismatch between the transmission line and the antenna at the input to the antenna. Figure 7.3-7 shows the time domain response of a UHF TV antenna system with a minor mismatch at the input of the antenna. Normally, such a mismatch is eliminated through the use of a fine matching network or a variable transmission line trans-

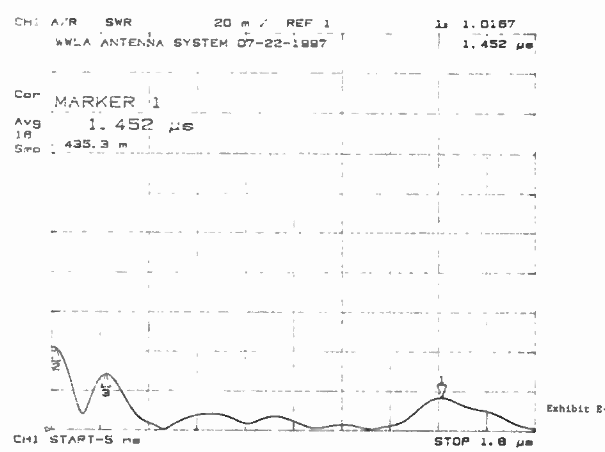


Figure 7.3-7(a). Time domain of TV antenna system with marker 1 at elbow complex at top of transmission lien before tuning.

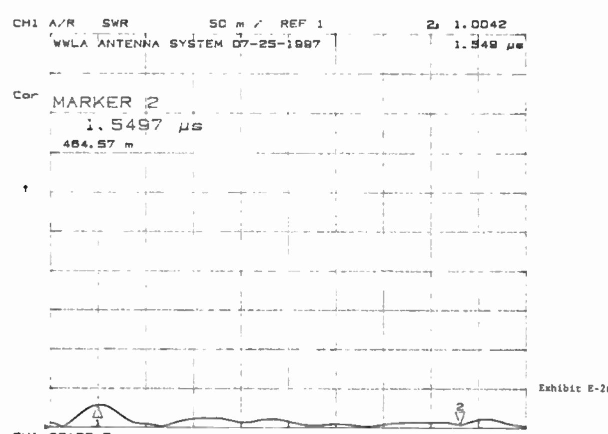


Figure 7.3-7(b). Time domain of TV antenna system with marker 1 on antenna input after system tuning.

former at that point in the system. As a rough rule of thumb, UHF antennas should be tuned for a maximum VSWR of 1.03 for the 6 MHz span of frequencies in the desired channel. To find problems with the transmission line system, the analysis should be made over a much broader range of frequencies. The greater the range of frequencies and the larger the number of points measured, the more detailed the information will be concerning the distance to faults. For one popular system, measurements are taken at 1601 points permitting details as small as the connections at each flange to be seen on a rigid coaxial line.

Waveguide Analysis

Due to its basic nature, a TDR will not work on a waveguide system because a waveguide does not work at low frequencies. Therefore, a TDR will always see a waveguide as an open circuit occurring at the transition from coaxial cable to the waveguide. However, in its simulated TDR mode, the network analyzer can determine the distance to any mismatch along a waveguide. For rectangular or truncated elliptical waveguides, this allows the location of mismatches to be determined. Tuning sections can then be installed (or tuning straps in the case of standard rectangular waveguide). Circular waveguide at UHF frequencies is normally not adjustable in the field other than the transitions from rectangular waveguide or coaxial cable.

Figure 7.3-8 shows the results of a properly tuned antenna system with waveguide. The network analyzer normally will store such information on a disk which permits the data to be displayed by simple printers.

FREQUENCY AND AURAL MODULATION MONITORING[1]

Various types of modulation monitors are discussed in this section, together with the FCC Rules with which modulation monitors and transmitter emissions must comply.

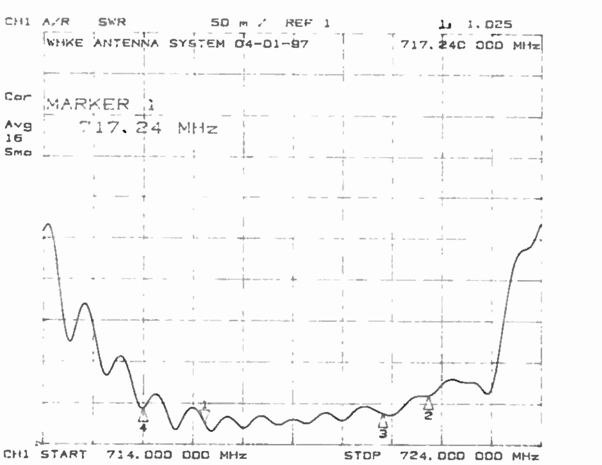


Figure 7.3-8. VSWR plot of tuned system. TV antenna with waveguide.

Modulation monitors are specialized, highly accurate and very low distortion instruments. They replace several more complicated, general purpose test instruments of similar accuracy as well as provide a convenient means for complying with FCC Rules.

Broadcast station monitors fall into two categories: frequency monitors and modulation monitors. These two functions are sometimes combined into one unit and sometimes packaged as separate units.

Broadcast aural modulation monitors are also classified according to the type of modulation to be monitored:

- AM monaural
- AM stereo
- FM monaural (and Subsidiary Communications Authority (SCA) as needed)
- FM stereo (and SCA as needed)
- TV monaural
- TV stereo (and multichannel sound)

Frequently, monaural and stereophonic functions are combined in one instrument for the convenience of the user.

WHY MONITOR THE TRANSMITTER

The three major reasons for using frequency and aural modulation monitors at a broadcast station are, described in the following sections.

Coverage

Because coverage of the broadcast service area is enhanced by a high modulation level, it is desirable to maintain the modulation level at the maximum legal limit in order to maintain good coverage and improve signal-to-noise ratio (SNR) for the audience.

Proof of Performance

By using the monitor as a precision test instrument, the station engineer can perform proof-of-performance measurements to ensure that the transmitter is within the modulation and frequency technical specifications of the manufacturer and the operating requirements of the FCC. The FCC no longer requires a regular "proof of performance." However, stations are still required to meet certain technical requirements.

Most aural modulation monitors have built-in facilities to measure baseband audio frequency response and signal-to-noise ratio. Outputs are provided for total harmonic distortion measurements by an external distortion analyzer. Modulation monitors with stereo and SCA functions can measure stereo separation, subcarrier injection levels, crosstalk between service channels, and provide demodulated outputs for stereo and SCA channels. Aural broadcast modulation monitors are often comprised of several integrated pieces of test equipment to test many functions of aural transmitters.

Compliance with FCC Rules

The monitor enables the station engineers to operate the transmitter in accordance with FCC Rules regarding aural modulation levels and carrier frequency tolerances. In addition, Part 73.1590(a) requires proof-of-performance measurements of all main transmitters, except Class D non-commercial educational FM stations operating under 10 W, to be taken as follows:

- Upon initial installations
- Upon modifications of transmission facilities
- Installation of AM stereo
- Installation of FM SCA or stereo
- Installation of TV stereo or subcarrier
- Annually on AM stations
- When required by other, special provisions of the station license.

FCC RULES AFFECTING MONITORING

The FCC no longer specifies the type of aural modulation monitor or measuring equipment a broadcast station must use. Therefore, it is the responsibility of the station licensee to decide what monitoring equipment is needed to ensure that the station transmitter emissions comply with FCC frequency and modulation requirements.

Frequency Monitoring

Part 73.1540(a) of the FCC Rules requires that center frequencies of AM, FM and TV stations must be measured or determined as often as necessary to ensure that they are maintained within the tolerances stated in FCC Rule 73.1545 as follows:

AM stations: The carrier frequency for monophonic transmissions or the center frequency for stereophonic transmissions must be within ±20 Hz of the assigned frequency.

FM stations: The center frequency must be within ±2000 Hz of the assigned frequency (±3000 Hz for transmitters having a power output of 10 watts or less).

TV stations: The visual carrier frequency must be within ±1000 Hz of the assigned frequency. The aural carrier frequency must be 4.5 MHz ±1000 Hz above the actual visual carrier frequency.

Aural Modulation Limits

Part 73.1570 of the FCC Rules states that modulation percentage is to be maintained at the highest level consistent with good transmission quality and broadcast service, not to exceed the following limits:

AM stations: Modulation of the carrier must not exceed 100% on negative peaks of frequent recurrence, or 125% on positive peaks at any time. There are additional regulations for AM stereo and telemetry transmissions.

FM stations: Total modulation must not exceed 100% on peaks of frequent recurrence referenced to 75 kHz deviation. However, stereo stations simultaneously providing subsidiary communication services (SCA) on subcarriers may increase the total peak modulation 0.5% for each 1.0% of subcarrier injection modulation; but the total carrier modulation must not exceed 110%. If two or more SCA subcarriers are used in conjunction with the stereo channel, the maximum allowable peak deviation is ±82.5 kHz or 110%.

TV stations: In general, the total modulation of the aural carrier must not exceed ±25 kHz deviation (monaural) on peaks of frequent recurrence. Stations transmitting multiplexed subcarrier signals on the aural carrier must limit the modulation of the aural carrier by the arithmetic sum of the subcarrier(s) allowable deviation and the total modulation must not exceed ±75 kHz deviation.

Modulation requirements for stations transmitting aural subcarriers as part of encoded subscription programs are stated in the application for FCC approval and contained in FCC Rule 73.682(b).

Modulation requirements for Broadcast Television Systems Committee (BTSC) stereo sound are subject to the criteria set forth in FCC Rule 73.682(c) and FCC Office of Engineering and Technology (OET) Bulletin 60A. Also see Chapter 6.3, *Multichannel Television Sound.*

ESSENTIAL FEATURES OF FREQUENCY MONITORS OR METERS

The primary standard of frequency measurements is the standard frequency maintained by the National Institute of Standards and Technology (NIST), which was formerly the National Bureau of Standards (NBS), or their standard broadcast signals of stations WWV, WWVB. Frequency monitors or meters must be capable of accurately measuring and displaying the carrier frequency requirements. Recommended resolution for AM monitoring is 1 Hz while FM and TV monitors should provide a resolution of 10 Hz or better. These tolerances call for an extremely accurate and stable internal frequency standard with an aging rate of 1 part per million (ppm) per year or better which is traceable to NIST.

If the transmitter is to be monitored at some distance from the transmitter site, a built-in or external preselector is generally required to raise the input level for receiving the signal off-the-air.

The following features are highly desirable in a frequency monitor:

- Digitally tuned RF preselector for multiple off-air applications
- Digitally tuned RF preselector for multiple off-air applications combined with the aural modulation monitor
- An output to operate an alarm when preset frequency limits are exceeded
- Provision for calibrating the internal frequency standard against an NIST station or other highly accurate standard
- An output for automatic logging.

Peak Modulation Duration

Although the FCC's current rules contain no precise definition of the maximum allowable peak modulation duration, some manufacturers of modulation monitors are using the pre-1983 de facto rules as their design guidelines. The *de facto* rules allowed a 1 millisecond response time for the peak modulation indicators and the permissible over-modulation limit was 10 counts per minute.

ESSENTIAL FEATURES OF AN AURAL MODULATION MONITOR

At a minimum, all aural modulation monitors should have a quasi-peak reading modulation meter to give a direct indication of modulation percentage. Meter accuracy should be ±4% or better. Other desirable features common to all modulation monitor are:

- Peak indicators, accurate to ±2% or better, to indicate when maximum positive and negative modulation peaks are occurring
- Adjustable peak indicator trigger points to indicate when modulation peaks exceed preset levels
- Adjustable peak modulation duration detecting circuit in the event the FCC and the broadcast community agree upon the maximum allowable peak duration
- An internal modulation level calibrator to check the accuracy of the modulation meter and peak flashers and a means for recalibrating the meter and peak flasher circuits.
- An output to operate an over-modulation alarm or a built-in alarm
- An output to operate an external alarm or a built-in alarm for when the modulation drops below a certain level (10%) for a specified period of time
- Outputs for a remote meter and peak flasher.

AM Monitors

AM monaural modulation monitors should possess the features described in the preceding paragraph. If an RF preselector is used for off-air monitoring with the monitor, its sensitivity should be approximately 100 μV for a 35 dB signal-to-noise-ratio (SNR) and 1 mV for a 50 dB SNR. Selectivity should be at least −40 dB at ±40 kHz, and image rejection should be at least 50 dB. The RF preselector must be very linear to avoid causing erroneous readings on the modulation monitor.

It is also desirable that the AM modulation monitor be equipped with NRSC deemphasis circuits so that the audio output of the monitor matches that of the AM transmitter modified with the NRSC audio response characteristics.

AM stereo monitors should have the following additional features:

- L + R and L − R channel decoding and outputs
- L and R channel separation measurement capability
- L and R channel SNR measurement capability
- L and R channel frequency response measurement capability
- Pilot carrier injection level measurement capability
- Channel crosstalk measurement capability
- Signal output for distortion measurements.

FM Monitors

FM monitors should have at least a 70 dB signal-to-noise ratio. The discriminator must have a distortion figure of 0.1% or better and a baseband frequency response of at least 25 Hz to 100 kHz, so that it can pass and accurately measure an SCA channel up to 92 kHz. FM monitors often consist of both frequency and modulation monitors in one package.

If an RF preselector is used for off-air monitoring, its intermediate frequency (IF) amplifier should have a linear phase response curve, yet be narrow enough to reject adjacent channels. A built-in multipath detector is highly desirable to help minimize multipath interference. The FM monitor should be equipped to measure synchronous and non-synchronous AM noise of the FM carrier.

FM stereo monitors should be able to measure:

- The L + R channel level (30 Hz to 15 kHz)
- The L − R channel level (23 to 53 kHz)
- The 19 kHz pilot injection level
- The 38 kHz subcarrier level
- Crosstalk between main channel and subcarriers
- Separation of left and right channels (up to 60 dB).

If one or more SCA subcarriers are transmitted, an SCA monitor should be used in addition to the FM stereo modulation monitor. The SCA monitor is usually an add-on to the main unit, which takes a composite feed from the demodulated output of the baseband monitor. Some manufacturers offer an optional RF and baseband demodulator, so that the SCA monitor can be used independently from the main monitor. The capability for user selection of SCA frequencies is important for future expansion of SCA service.

The SCA monitor should be able to measure:

- Modulation percentage
- SCA injection level on the composite signal
- Signal-to-noise ratio
- Crosstalk.

The SCA modulation measurement should be selectable for ±4 kHz or ±6 kHz as the level for a meter indication of 100%. FM monitors often include both modulation and frequency monitoring in one package.

Analog TV Monitors

The features for analog TV monaural and stereo monitors are similar to those for FM monitors. When transmitting BTSC multichannel sound, the operator should be able to monitor the main channel and stereo channel as well as SAP and PRO channels (if they are utilized). Monitoring the modulation level of the BTSC signal is extremely important for achieving good stereo separation.

BTSC Stereo Separation and Modulation Accuracy

Stereo separation in the BTSC format is sensitive to gain and phase errors in the transmission path. This is because the L + R and L − R signals are treated differently. In particular, L − R is companded while L + R is simply preemphasized and deemphasized. The L − R and L + R signals must arrive at the receiver's decoder matrix, which yields L and R, with very small errors in gain and phase across the aural baseband from 50 Hz to 50 kHz. Figure 7.3-9 shows how stereophonic separation is affected by gain and phase errors in the L − R signal relative to the L + R signal at the input of the final matrix.

Subjective tests have shown that an average listener begins to "perceive" a loss in the spatial character of stereophonic music material when the separation drops below 18 dB. A separation of 15 dB ± 3 dB is considered "adequate" by the average listener.

Although the subjective effects of separation depend on the spectral distribution and other aspects of the audio material, it appears that a good engineering objective for the entire system is for the separation to exceed 20 dB in the mid-range, decreasing somewhat at frequencies above 8 kHz. Figure 7.3-9 shows that a separation of 20 dB requires a gain error smaller than 1 dB, and a phase error of less than 10°. The BTSC standards require that the separation of the radiated signal exceed 30 dB in the mid-band from 100 Hz to 8 kHz, but that is may decrease at low frequencies to 26 dB at 50 Hz, and to 20 dB at 14 kHz. This requires that the gain and phase errors in the mid-band be less than 0.3 dB and 3.0°, respectively.

The total modulation level accuracy in BTSC stereo is more critical than in FM stereo radio broadcasting to produce acceptable stereo separation. Because the L + R and L − R signal paths in BTSC are processed differently, a small change of modulation level in the BTSC system will affect the stereo separation. This is because the amplitude and phase relationship between the L + R and L − R channel is altered. If the total modulation level of the BTSC system is not maintained accurately, the dbx decoder in the receiver will see an incorrect RMS level and reproduce an L − R signal with altered amplitude and phase. That is, if an incorrect L − R signal is fed to the decoding matrix, the consequence will be poor stereo separation. Because the RMS level to the input of the decoder is directly proportional to the total modulation level, the total modulation level in the BTSC transmitter must, therefore, be accurately monitored in order to maintain good stereo separation and high quality audio performance.

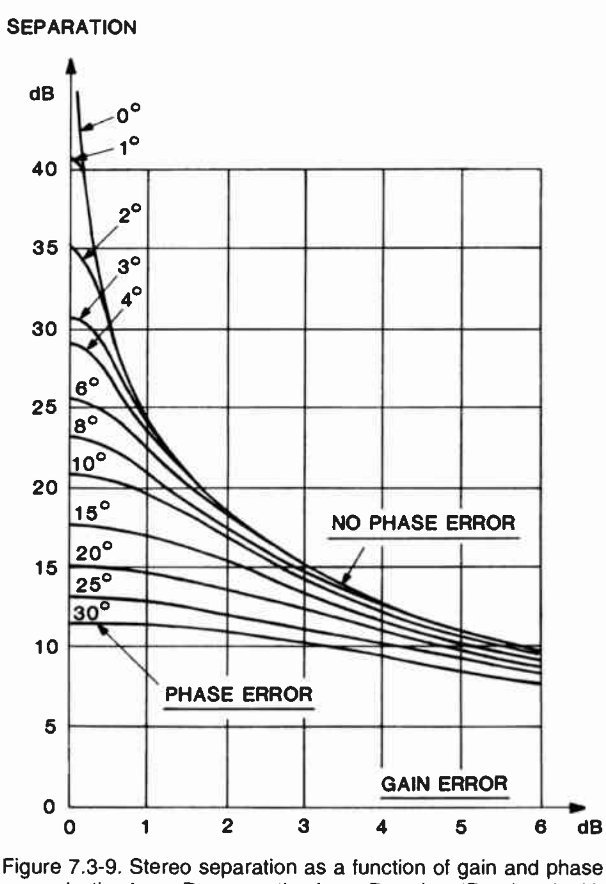


Figure 7.3-9. Stereo separation as a function of gain and phase errors in the L − R versus the L + R paths. (Reprinted with permission from TFT, TV Aural Proof of Performance Guide.)

ON-SITE MONITORING TECHNIQUES

In a studio-transmitter collocated operation, the monitor is normally connected directly, or through an attenuator, to an RF sampling point of the transmission line feeding the antenna via a directional coupler. It is important to know the RF voltage level of this sampling point so that it meets the input requirement of the monitor.

AM Monitoring

Most AM frequency monitors display carrier frequency error rather than actual carrier frequency. Typically, a front-panel digital display indicates the carrier deviation in Hz from its assigned frequency, and a lighted "+" or "−" indicates whether the carrier is above or below the carrier frequency.

The modulation percentage of a monaural AM carrier is normally displayed directly on a front-panel meter.

Peak flashers, used on some monitors, are intended to catch fast transients and peaks that the meter cannot respond to. There may be one flasher to indicate maximum allowable negative peaks (100%) and another to indicate maximum allowable positive peaks (125%). The monitor may also have an adjustable peak flasher which can be preset by means of digital switches so that it flashes when the modulation percentage exceeds the preset switch value.

For monitoring modulation of an AM stereo transmitter, an AM stereo monitor or a stereo monitor plus a compatible AM modulation monitor is required. The equipment should permit the operator to simultaneously read the modulation percentage on both left and right channels and to measure separation between channels and crosstalk between the main channel and subchannels.

FM Monitoring

A typical FM frequency monitor digitally displays carrier frequency error in Hz (rather than display actual carrier frequency) and indicates whether the carrier is above or below its assigned frequency. The RF input level must be adjusted so that sufficient signal is available for measurement but not so high that overloading occurs. In addition to monitoring the FM stereo transmitter carrier frequency, it should be possible to monitor the stereo pilot carrier frequency as well.

On a typical monaural FM monitor, the modulation percentage is normally displayed on a front-panel meter. Generally, either positive, negative or combined modulation peaks can be selected for monitoring. Some monitors also provide peak flashers to indicate modulation peaks that exceed a preset percentage. The RF input must be adjusted to the correct level, as described in the instruction manual, before accurate readings can be taken.

For monitoring a stereo FM transmitter, a modulation monitor with a compatible stereo monitor must be used. Left channel, right channel and total modulation are generally read on front-panel meters. With the typical monitor, these meters can also be used to measure separation between the left and right channels as well as crosstalk between the main channel and subchannels.

SCA monitors are usually accessories to the baseband monitor. They are either fed from the FM composite signal or from the RF carrier through a separate FM demodulator.

Analog TV Monitoring

The typical TV frequency monitor will provide separate displays of the visual carrier, aural carrier and intercarrier frequency errors (rather than actual carrier frequency). The RF input to the monitor should be adjusted to the correct level as described in the instruction manual before measurements are made.

Aural modulation percentage can be read from a front-panel meter. If the TV transmitter is also transmitting BTSC stereo sound, a monitor with a stereo decoder for left and right channels must be used. If a SAP channel is employed, a SAP monitor should also be used to monitor modulation and frequency.

Due to the critical relationship in BTSC stereo separation between the recovery of the companded L − R channel and the modulation level of the transmitter described previously, it is essential to maintain accurate modulation levels when broadcasting multichannel television sound. One convenient method of achieving this objective is to monitor the pilot carrier injection level, which has a constant level. After calibration of modulation levels for maximum stereo separation, the pilot carrier injection level is the best reference for maintaining maximum stereo performance.

OFF-SITE MONITORING TECHNIQUES

For remote, off-air monitoring, the monitor should incorporate a built-in preselector (RF amplifier) or an external RF preselector connected to an outdoor antenna. Some preselectors are capable of measuring and indicating carrier frequency error. Such preselectors, when combined with a modulation monitor, provide a compact complement of monitoring equipment.

Off-site monitoring techniques are generally the same as on-site techniques. RF input level to the preselector or monitor must be carefully adjusted, as described in the manufacturers' instruction manuals. It is even more important to know the RF level so that it does not overload the RF preselector and create intermodulation products. Some monitors are equipped with a multipath detector which enables the user to rotate the receiving antenna for minimum multipath interference.

CALIBRATION AND MAINTENANCE

Frequency monitors should be calibrated periodically to ensure accurate measurement of transmitter carrier frequency error. The time between calibrations depends on the frequency of the monitor's internal standard, which in turn depends on the operating frequency of the transmitter. For a monitor having an internal crystal standard that has been correctly calibrated once and that has an aging rate of 1 ppm per year, the following calibration schedule is recommended:

Transmitter Frequency	Calibration Interval
AM band	Every 12 months
FM band	Every 12 months
Low VHF Analog TV band (Ch 2–6)	Every 12 months
High VHF Analog TV band (Ch 7–13)	Every 6 months
UHF Analog TV band	Every 3 months

After the first few calibrations, the interval may be lengthened if the drift observed in the first few checks warrants it.

There are a number of ways the internal frequency standard in a monitor can be calibrated:

- Use a high quality frequency counter having a resolution of 1 Hz display and a time base that is calibrated against a secondary standard having a accuracy of at least 1×10^{-8}
- If the frequency display can be operated in a general-purpose counter mode, a frequency standard of higher accuracy than the monitor's internal standard can be used with the counter to calibrate the internal standard. The external standard is connected to the counter input, and the internal standard (which furnishes the time base for the counter) is adjusted so that the counter displays the exact frequency of the external standard
- If a receiver capable of receiving one of the NIST stations (WWV at exactly 5, 10 and 15 MHz) is available and if the monitor has an output from its internal standard that is an exact subharmonic of the

frequency of the NIST station being received, the monitor internal standard can be calibrated by adjusting its frequency to zero beat with the NIST station frequency.

Aural Modulation Monitor Calibration

The aural modulation meter and peak flashers should be calibrated regularly. Most monitors have built-in calibrators, so that meter and flasher accuracy can be checked by simply pressing front-panel switches and observing the peak flashers and the meter reading. If the reading is in error, a simple adjustment usually corrects the error.

If the built-in modulation calibrator is of the frequency marker type for establishing the $\pm 100\%$ peak modulation level, and the frequency markers are generated from single crystal sources, the accuracy can usually be maintained within $\pm 1\%$. If this type of monitor is being used, it is not necessary to calibrate the FM modulator using a Bessel null method.

If an FM modulation monitor has no internal calibrator, the monitor must be calibrated against a laboratory standard or by means of a Bessel null measurement using a spectrum analyzer and a precision audio frequency generator.

If an AM modulation monitor has no internal calibrator, the monitor can be calibrated using an RF generator. The generator must be capable of very low distortion amplitude modulation, and the level of this modulation should be accurately observed by using a high quality oscilloscope.

An AM monitor is usually calibrated by comparing the peak amplitude of the waveform against an amplitude reference established by an oscilloscope with high linearity. A digitally generated reference is frequently used as a built-in calibration standard as in an FM modulation monitor.

Maintenance

A broadcast station monitor should be maintained in the same way as other precision laboratory instruments. If should be calibrated regularly as described in the preceding paragraphs. Manufacturers also offer this type of service to their customers.

DIGITAL TELEVISION (DTV) MEASUREMENT ISSUES[2]

For analog signals, some transmission impairments are tolerable because the effect at the receiver is often negligible, even for some fairly significant faults.[3] With DTV, however, an improperly adjusted transmitter could mean the loss of viewers in the Grade B coverage area (or worse). DTV reception—of course—does not degrade gracefully, it simply disappears. Attention to several parameters are required for satisfactory operation of the 8VSB modulation system used by DTV. First, there is the basic FCC requirement against creating interference to other over-the-air services. To verify that there is no leakage into adjacent channels, out-of-band emission testing is required. Second, while for NTSC analog TV, there is concern with SNR performance, in the DTV case the desired-to-undesired signal ratio (D/U) is measured.

In DTV, flat frequency response across the channel passband is required comparable to that required by analog transmitters. A properly aligned DTV transmitter exhibits many of the same characteristics as a properly aligned analog unit, including flat frequency response and group delay, with no leakage into adjacent channels. Where analog TV is concerned, group delay problems result in chromanance/luminance delay, which can degrade the displayed picture but still leave it viewable. Group delay problems in DTV transmitters, however, result in intersymbol interference (ISI) and a rise in the bit error rate (BER), possibly causing the receiver to drop in and out of lock. Even low levels of ISI may cause receivers operating near the edge of the cliff to lose the picture completely. Amplitude and phase errors may cause similar problems, again resulting in reduced viewer coverage.

Eye patterns and BER have become well-known parameters of digital signal measurement, although they may not always be the best parameters to monitor for 8VSB transmission. The constellation diagram and modulation error ratio, on the other hand, provide insight into the overall system health. RF constellations are displayed on the (I) in phase and (Q) quadrature components of the received RF signal, commonly displayed on the X and Y axes of the constellation diagram, respectively. Constellations of tight vertical dot patterns with no slanting or bending indicate proper operation, as illustrated in Figure 7.3-10(a). Since 8VSB levels are the in-phase signal they are displayed left to right.

An 8VSB signal is a single sideband signal with a pilot carrier added. In a single sideband signal, phase does not remain constant. Therefore, the constellation

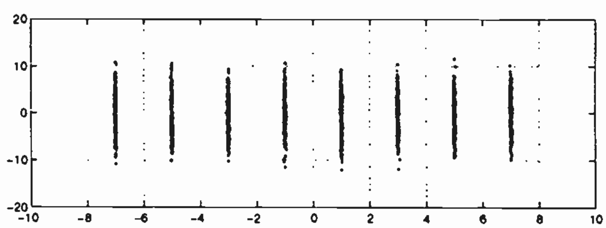
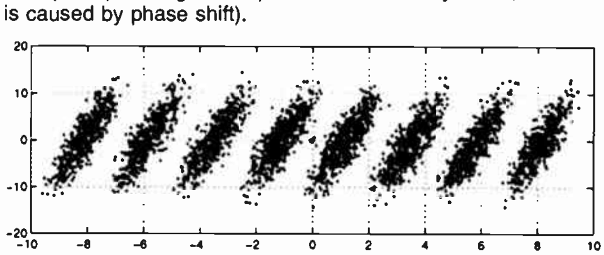


Figure 7.3-10(a) and (b). 8 VSB constellation diagram: (a) a near-perfect condition, (b) constellation diagram with noise and phase shift (the spreading of the pattern is caused by noise; the slant is caused by phase shift).



points (dots) occur in a vertical pattern. As long as the dot pattern is vertical and the points form narrow lines of equal height, the signal is considered good and can be decoded.

Figure 7.3-10(b) shows an 8VSB signal that has noise and phase shift. Noise is indicated by the spreading of the dot pattern. Phase problems are indicated by the slant along the Q-axis.

While BER is a valid measurement for 8VSB, a better approach involves monitoring the modulation error ratio (MER), which usually will reveal problems before the BER is affected. In many cases, MER can provide enough warning time to correct problems that would result in an increase in the BER. MER provides an indication of how far the points in the constellation have migrated from the ideal. There can be considerable migration before boundary limits are exceeded. Degradation in the BER is only apparent when those limits have been exceeded.

Principle DTV System Parameters

The FCC, in its Memorandum Opinion and Order on Reconsideration of the *Sixth Report and Order* on DTV (February, 1998), specified the out-of-band emissions mask for terrestrial transmission. The order requires that transmitter out-of-band emissions be attenuated consistent with the following:

- In the first 500 kHz from the authorized channel edge, emissions must be attenuated no less than 47 dB below the average transmitted power
- At more than 6 MHz from the channel edge, emissions must be attenuated no less than 110 dB below the average transmitted power
- At any frequency between 0.5 and 6 MHz from the channel edge, emissions must be attenuated no less than the value determined by the following formula, which is based on a measurement bandwidth of 500 kHz: attenuation in dB $= -11.45(\Delta f + 3.6)$, where $\Delta f =$ the frequency difference in MHz from the edge of the channel.

When the interference level of the spectral distribution as defined here is not low enough for the operational conditions of a particular application, other spectral distributions should be considered.[4]

In-band Signal Characterization

The quality of the in-band emitted signal can be specified and measured by determining the departure from the 100% eye opening. This departure, or error, has three components:

- Circuit noise (white noise)
- Intermodulation noise caused by various non linearities
- Intersymbol interference

The combination of all of these effects can be specified and measured by an error vector magnitude. This measurement is described in Reference 4.

Power Specification and Measurement

Conventional NTSC broadcast service allows a power variation ranging between 80% and 110% of authorized power. These values correspond to -0.97 dB and $+0.41$ dB, respectively. Because of the cliff effect at the fringes of the service coverage area for a DTV signal, the allowable lower power value will have a direct impact on the DTV reception threshold. A reduction of 0.97 dB in transmitted power will change the DTV threshold of 14.9 dB (which has been determined to yield a 3×10^{-6} error rate) to 15.87 dB, or approximately a one mile (1.6 km) reduction in coverage distance from the transmitter. Therefore, the average power of the DTV transmitted signal is of significant importance.

The ATSC recommends a lower allowed power value of 95% of authorized power and an upper allowed power value of 105% of authorized power.

OTHER MEASUREMENT CONSIDERATIONS

It is impossible to cover all of the measurement requirements for a particular installation in this chapter. Transmitter manufacturers usually provide extensive documentation on the tests and measurements recommended and/or required for their products. In-factory or on-site instruction is also available from many firms. Because the purchase of a transmitter or antenna system is a sizeable investment, the instrumentation and educational requirements of upgrading an RF plant should be considered before the purchase of equipment, not afterwards.

REFERENCES

1. Wu, Joe. "Frequency and Aural Modulation Monitoring," *NAB Engineering Handbook,* 8th. edition pp. 609–614, 1992.
2. Whitaker, Jerry, C. "Video Measurement Techniques," *DTV: The Revolution in Electronic Imaging,* McGraw-Hill, New York, NY, pp. 469–471, 1998.
3. Reed-Nickerson, Linc., "Understanding and Testing the 8VSB Signal," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, pp. 62–69, November 1997.
4. ATSC, "Transmission Measurement and Compliance for Digital Television," Advanced Television Systems Committee, Washington, DC, Doc. A/64, November 17, 1997.

Section 8: Ancillary Broadcast Systems

Introduction by Al Resnick, ABC, New York, NY

Within the last decade, the radio industry has seen explosive growth in studio and transmission facility construction. Stations and studios have been bought, sold and modified at a record pace. With the dawn of DTV, a similar construction boom is likely for television.

Building a new studio or transmission facility is a major undertaking. Many engineers dream about becoming involved in such a project. The difference between a project being a dream or a nightmare often rests in the preparations and planning.

Studio design and construction is the most visible and is—arguably—the most important task of the broadcast engineer. This work involves not only the physical space of the facility, but also the infrastructure, which includes everything from the ac power and grounding subsystems to the broadcast equipment itself. The provision of a proper environment is fundamental to the safe and reliable operation of a station. Consolidation of duties has brought us back to the drawing board to consider more efficient floor plans that maximize the use of space and time.

The successful planning and construction of a broadcast facility is always time well spent and can deliver benefits as quickly as the start of construction as well as efficiency and flexibility of the physical plant long into the future. Learn how to ensure a successful project for your station in Frank Rees, Jr.'s *Broadcast Facility Planning and Construction* chapter.

Much of today's broadcast equipment is computer based. Most computer-based equipment is seldom ready to operate instantly upon application of power. The facilities planner must evaluate options to keep modern equipment operating through those times when the local electrical supply has problems. Learn how to protect your ac power system from damage as *Handbook* Editor-in-Chief, Jerry Whitaker provides a guide to the wide range of power conditioning and isolation equipment in his *AC Power Conditioning* chapter.

W. E. Dewitt's comprehensive *Facility Grounding Practices* chapter defines grounding and lists the reasons and methods for proper grounding, discusses equipment grounding, system grounding and earth connections, and applies recommended facility grounding practice to the safe design and construction of electronics installations.

As the section concludes, Jerry Whitaker weighs in once again with valuable information for maximizing reliable operation when faced with emergency situations in his *Standby Power Systems* chapter.

8.1
BROADCAST FACILITY PLANNING AND CONSTRUCTION

FRANK REES, JR.
REES ASSOCIATES, DALLAS, TX

INTRODUCTION

Planning and construction of a broadcast facility must combine today's state-of-the-art excellence with the flexibility and unknown elements of tomorrow. With the correct approach, this is not a daunting task. By understanding short-term and long-term needs, broadcast facility owners can create a building and infrastructure that is geared to sustain a station throughout the next half-century.

A STRATEGY FOR SUCCESS

No one understands more clearly the sweeping technological changes impacting the broadcast industry than the individuals that are currently limited by their physical environment as a result of its changes. The impact of these technological changes can be staggering. Considering some broadcast facilities may not have the capacity to handle current needs, the facility cannot begin to support changes anticipated within the next few years. When launching into the planning and construction of a broadcast facility, meeting these current and future needs requires careful consideration and a well defined strategic process.

When detailing a successful planning and construction strategy for a broadcast facility, it is imperative that owners keep several important elements in mind. Flexibility is primary and function dictates design needs. The final design must accommodate construction and operational efficiencies. Finally, choose the planning and construction team wisely. When planning for design and construction of a broadcast facility, the elements of flexibility, efficiency and a quality team are paramount.

Flexibility is the linchpin of planning. The owner must plan for current and future needs. Planning takes knowledge of the industry and an intuition for the future. A good plan must allow for technology and equipment that currently does not even exist.

There must be efficiency in design. Efficiency in the design of a broadcast facility reduces the initial cost, lowers the operational cost and allows the staff to be more productive in their everyday tasks. To achieve this efficiency, each team member has to be focused on specific goals to form a concise overall plan. Focusing the many different disciplines required to design a broadcast facility requires a strong team leader who is knowledgeable of the numerous disciplines, as well as the owner's needs, as they relate to a broadcast facility.

The planning and construction team is three-pronged, it includes the owner, architects/engineers/consultants and contractor. By combining these three integral team players, a multidisciplinary effort results in a clear, masterful approach to the planning and construction of a broadcast facility.

Success of the project is highly contingent upon the degree of the owner's participation and willingness to provide the information vital to the creation of building plans. The owner must be involved in all phases of the project in a formal, aggressive way if the project is to successfully come to fruition. After all, the owner's vision is what gives substance to the project.

The architects are the experts who guide and lead the team from the planning stages through construction. From there they move toward completing a building that masters design, flexibility and function and provides an infrastructure to support these elements on a short and long-term basis. The experience of the architect cannot be overlooked. A poor choice can cost time and dollars now and in the future.

The architect's design team also includes professionals uniquely qualified to design the various engineering systems of a broadcast facility. This roster of professionals includes electrical, mechanical, structural and acoustical engineers and broadcast system engineers and integrators, as well as other consultants.

The same philosophy of selecting a qualified architect applies to the contractor. It is important that the construction professional is involved from the planning phase through the construction process of a broadcast facility.

In summary, flexibility, efficiency of design and a quality team are the key components of the strategic planning process. These elements should support the project goals and objectives and create the framework to successfully plan and construct a broadcast facility.

THE PLANNING, DESIGN AND CONSTRUCTION PHASES

Development of a broadcast facility is like a journey—it requires careful planning and a detailed road map.

The Planning Process

The planning process should be developed prior to the design of the actual structure, to obtain a comprehensive overview of the proposed facility. This process serves as the road map, which analyzes the entire trip from the beginning, and defines the building's size, cost and time of completion by studying every aspect of the project.

During the development of the facility plan, the owner typically has two primary goals:

- Improve the functional quality of the facility and thus the value of the completed project by involving everyone (station management and staff as well as architects, builders and systems engineers) in the planning and design of the facility
- Minimize the cost and time required to design and construct the facility.

Following is an outline of ten suggested components that are the basis for a meaningful process for the planning and construction of a broadcast facility:

- Project issues are those broad stroke components that define the overall operational, business and facility requirements for the success of the facility. These issues guide the project goals and can also range from the technical to the aesthetic expectations for the project
- If applicable, an existing facility analysis is completed to determine if it is appropriate to utilize the existing facility for the future or if there is a need for new construction
- Planning criteria includes those principles that direct the design team to adhere to basic design guidelines including Americans with Disabilities Act (ADA) requirements, energy efficiency, function driven planning and future expansion strategies. Adaptability and visually pleasing qualities can both be a part of the overall project requirements
- A functional space program outlining the requirements of each portion of the broadcast facility is developed. This lists the size and quantity of each functional space within the facility. This functional description is most useful when all aspects of the space are listed including furniture, equipment, unique lighting and electrical and mechanical systems. The program can also define where each space belongs in the overall scheme
- A building systems description and quality statement which outline the method of construction and level of quality of all portions of the broadcast facility must be developed as a basis for the probable budget statement
- A project development schedule which outlines the time necessary for all planning, design, corporate approval, construction, broadcast equipment installation, move in/startup and other activities necessary to achieve the target on-air date for the facility
- Facility concepts and site analyses which outline the size and location of all portions of the broadcast facility including future expansion
- A probable budget statement which outlines the costs of all portions of the broadcast facility including construction, furnishings, technical equipment, development costs and other related expenses. A guaranteed maximum cost for the building could be established if the level of contractor involvement allows for a clear understanding of the scope of the work
- A cash flow schedule indicating total and monthly project cash flow requirements is incorporated into the development schedule
- Design studies to show the physical appearance of the completed project.

The Design Process

Once the planning process has been completed, the subsequent phases of design and construction can begin. The design and construction process can easily be divided into five significant phases:

- Schematic design
- Design development
- Construction documents
- Bid negotiation/pricing
- Construction.

Each of these phases works independently, and together, to bring about the best end product.

All the best planning for design and construction, however, cannot succeed without giving proper attention to the *construction process*. It is important to choose a construction process approach that is best suited to the project and a construction professional whose experience and expertise can support building a highly technical facility.

Preparing your Building for Digital Television

Whether renovating an existing station or building a new television facility, today's digital television facility needs will be different from yesterday's analog plant. As an organization plans for the project, whether it involves one channel or multiple channels, proper planning for the organization's facility, studio and transmission needs will ensure success.

The Technical Core

Heating, ventilation and air conditioning systems (HVAC) in the technical areas require very reliable temperature and humidity control as well as redundancy to keep the sophisticated electronic equipment from shutting down due to excessive temperature conditions (above 85° F). The HVAC systems must be designed to ensure continuous uninterrupted service. The use of self contained *computer room air conditioning* (CRAC) units with built in redundancy (dual compressors or with chilled water with direct expansion backup) is common in the electronic rack room and related technical operations areas. Proper humidity control can be easily achieved with these units when used in conjunction with proper vapor barrier placement. Proper unit location and acoustical design is required to address acoustical concerns in critical listening technical and production spaces.

Reliable power is an extremely critical component in digital radio or television stations. Two separate sources of primary electrical service are often obtainable from the local power company via power from two substations or two separate power grids. However, the ability to quickly switch from the main to an alternate power source is critical. The power company will provide information on power characteristics for a given location. If two power sources are not available then a standby generator may be required. The generator should be sized to carry the critical loads of the facility, including HVAC systems for the technical and transmitter areas. An *uninterruptible power supply* (UPS) will also be required to bridge the time gap when power is switched over and the standby generator comes on-line or to give personnel time to power down computer based systems with volatile memory.

Electrical service for a facility can easily be designed to separate the load into several branches. This will allow technical power to be isolated from mechanical equipment, studio lighting and normal building power to minimize the cost to provide clean power to the critical technical areas.

A properly designed technical *grounding system* is critical to station operations. A *lightning protection system* is desirable and will minimize damage to sensitive electronic components and pay for itself in the event of a lightning strike.

The design of the *technical core* is also critical for digital broadcast stations. The technical spaces are larger than in the past although individual pieces of equipment may be smaller. With the advent of server technology more equipment requiring more space will be the norm. Much equipment requires the use of 36 in. deep equipment racks. This increase in depth along with considerations for handicap access, required under the ADA, requires more space in the technical areas than ever before. The additional equipment required to broadcast digital television as compared to analog, impacts space requirements along with the requirement of multiple channels transmitting from one facility. The additional rack and equipment space required for multiple channels will require more space as well as more HVAC and electrical systems support. Building systems to support technical core will be much the same as they have been in the past but will require more flexibility.

Technical core spaces will still require a minimum of 18 in. access flooring for cable requirements. Overhead raceway systems could also be an alternative in the distribution of cable systems throughout the building. Similar systems will be required for newsroom operations and editing systems. Typical edit suites will also require the continued use of access flooring.

The key to successful planning and construction of the next generation station will be flexibility. It is imperative that the buildings of tomorrow be more flexible and communicate more easily within itself for the user to enhance its capabilities. Fewer walls and open office modular workstations will become the norm instead of the hard wall offices of the past.

Master Control

Stations that carry multiple DTV channels will impact master control as well as electronic rack room and transmission areas. Monitoring, switching and playback will have varying degrees of impact depending on how many channels are carried and the types of digital video systems utilized.

Studio Control and Post Production Suites

Equipment required for digital production switchers and editing systems required in control rooms and edit suites requires less space than the large bulky consoles of the past. Monitor walls must be designed for visual comfort of personnel within the room with rear access for service provided. The main design concern is the space required for personnel to function properly.

Studios

Studios for DTV will need to be designed to accommodate the 16:9 aspect ratio. The decreasing need for studio space due to virtual sets and digital technology will impact total space requirements of the television station. Floor space will be greater to accommodate the wider shot, however, the required height of the lighting grid may be less in the long-term.

Newsroom

The ability to edit video from the desktop using digital networked systems may reduce the number of traditional viewing and editing stations previously required with analog systems. Reporters and others needing to view or edit video may be able to do so using desktop computers.

Additional Considerations

The distribution of broadband signal systems throughout the station will require that every desktop be connected to the servers and video systems residing in master control and elsewhere. This networked technology will allow team editing and instant communication regardless of the physical layout of the station or individual departments.

Sophisticated computer use will require dedicated computer training rooms to ensure that staff may stay abreast of new software. The ability to broadcast multiple channel CD quality sound and multiple channel standard and high definition video will enable the broadcaster to transmit multiple programs while serving other businesses. This requires operational and facility planning to be accomplished in an orderly systematic manner.

THE CONSTRUCTION PROCESS

Within the construction phase, there are three primary ways to approach the construction process: the *conventional approach* (*competitive bid*), *design/build approach* and *negotiated approach*.

Conventional Approach (Competitive Bid)

Once the construction documents are completed, the bid phase begins. Drawings are issued to a select list of construction professionals who then competitively bid on the project. Typically, the contractor with the lowest bid is awarded the contract. A contract is executed between the owner and the contractor.

Design/Build Approach

Contractor and architect, as a team, design and construct the project. A single contract is executed between the owner and the design team. The design team then provides a complete turn key project for the owner.

Negotiated Approach

The negotiated approach involves preselecting a group of contracting firms for interviews. The construction firm may be selected on previous experience, personnel and/or cost. Just like the architectural firm, the contractor is selected based on qualifications. The construction firm should be included from the beginning of the project through construction of the building.

For broadcast facilities, the negotiated approach is the most frequently used construction process, followed by the design/build approach and the competitive bid approach.

Selecting a Contractor

The construction professional's expertise is an integral part of the project's success and ranges from scheduling, keeping within budgeted dollars and material considerations to constructability. This key player must also have a good knowledge of technical facilities and understand the importance of flexibility for the future.

The determination of who will be retained to construct the new or renovated building is one of the most crucial decisions a broadcast facility owner makes when planning and constructing a broadcast facility. There can be serious trouble from the first day if a contractor or construction professional with the proper qualifications is not selected for the project.

When selecting a general contractor, there are four main areas to investigate:

- Evaluate the history of the firm. A consistent management philosophy has most likely been established if the firm has been in business for at least ten years
- Evaluate the reputation of the firm, its leadership and subcontractors. Does the reputation of the firm and its President coincide with your company standards? Does the contractor have a reputation for litigation with clients and subcontractors?
- Evaluate the firm's financial stability. Select a firm with a net worth of at least half the projected costs of the planned project
- Evaluate the firm's past project record, especially technical facilities. Request a list from the past five years, sorted by year, of total construction costs the firm has completed. Determine the bonding limit of the contracting firm and make sure this dollar amount is at least twice the construction budget. It will guarantee the performance of the contract and the insurance company can also provide valuable information on the company. Ask for a listing of at least three of the company's past projects. Detailed information about each project should include the project's location, construction budget, year built, and contact name at the facility. If a contractor does not have experience with highly technical facilities, that firm is not qualified to construct a broadcast building.

The relationship with the contractor is interactive and long-term. Before the contractor has been selected, request a detailed description of how the firm intends to approach the construction of the broadcast facility. It is important to secure information on how the firm intends to keep the team informed and how they will provide the team with an understanding of what is transpiring during the construction process.

CONCLUSION

When embarking on the venture of planning and constructing a broadcast facility, the key to success begins with clearly defined goals and objectives that incorporate flexibility and efficiency in every aspect of the project. Additionally, achieving short and long-term objectives can be met by using a tested and effectual design approach with a qualified team whose commitment, vision, experience and expertise can guide the planning and construction process through strategic phases. Overall, these elements blend to create a broadcast facility planning and construction mix that will bring the project to an impressive conclusion and create a facility that meets the operational needs of today and provides flexibility for tomorrow.

8.2
AC POWER CONDITIONING

JERRY C. WHITAKER, EDITOR-IN-CHIEF
MORGAN HILLS, CA

INTRODUCTION

Utility companies make a good-faith attempt to deliver clean, well regulated power to their customers. Most disturbances on the ac line are beyond the control of the utility company. Large load changes imposed by customers on a random basis, power factor (PF) correction switching, lightning and accident related system faults all combine to produce an environment in which tight control over ac power quality is difficult to maintain. Therefore, the responsibility for ensuring ac power quality must rest with the users of sensitive equipment—in this case, the broadcaster.

The selection of a protection method for a given facility is as much an economic question as it is a technical one. A wide range of power line conditioning and isolation equipment is available. A logical decision about how to proceed can be made only with accurate, documented data on the types of disturbances typically found on the ac power service to the facility. The protection equipment chosen must be matched to the problems that exist on the line. Using inexpensive basic protectors may not be much better than operating directly from the ac line. Conversely, the use of a sophisticated protector designed to shield the plant from every conceivable power disturbance may not be economically justifiable.

Purchasing transient suppression equipment is only one element in the selection equation. Consider the costs associated with site preparation, installation and maintenance. Also consider the operating efficiency of the system. Protection units that are placed in series with the load consume a certain amount of power and, therefore, generate heat. These considerations may not be significant, but they should be taken into account. Prepare a complete life-cycle cost analysis of the protection methods proposed. The study may reveal that the long-term operating expense of one system outweighs the lower purchase price of another.

The amount of money a facility manager is willing to spend on protection from utility company disturbances generally depends on the engineering budget and how much the plant has to lose. Spending $225,000 on system-wide protection for a major market television station is easily justified. At smaller operations, justification may not be so easy.

The Key Tolerance Envelope

The susceptibility of electronic equipment to failure because of disturbances on the ac power line has been studied by many organizations. The benchmark study was conducted by the Naval Facilities Engineering Command. The far-reaching program, directed from 1968–1978 by Lt. Thomas Key, identified three distinct categories of recurring disturbances on utility company power systems[1]. As shown in Table 8.2-1, it is not the magnitude of the voltage, but the duration of the disturbance, that determines the classification.

In the study, Key found that most data processing (DP) equipment failure caused by ac line disturbances occurred during bad weather, as shown in Table 8.2-2. According to a report on the findings, the incidence of thunderstorms in an area may be used to predict the number of failures. The type of power transmission system used by the utility company also was found to affect the number of disturbances observed on power company lines (see Table 8.2-3). For example, an analysis of utility system problems in Washington, DC, Norfolk, VA, and Charleston, SC, demonstrated that underground power distribution systems experienced one-third fewer failures than overhead lines in the same areas. Based on his research,

Table 8.2-1
Types of Voltage Disturbances Identified in the Key Report[1]

Parameter	Type 1	Type 2	Type 3
Definition	Transient and oscillatory overvoltage	Momentary undervoltage or overvoltage	Power outage
Causes	Lightning, power network switching, operation of other loads	Power system faults, large load changes, utility company equipment malfunctions	Power system faults, unacceptable load changes, utility equipment malfunctions
Threshold[1]	200-400% of rated rms voltage or higher (peak instantaneous above or below rated rms)	Below 80-85% and above 110% of rated rms voltage	Below 80-85% of rated rms voltage
Duration	Transients 0.5-200 (μs wide and oscillatory up to 16.7 ms at frequencies of 200 Hz to 5 kHz and higher	From 4-6 cycles, depending on the type of power system distribution equipment	From 2-60 s if correction is automatic; from 15 min to 4 hr if manual

[1] The approximate limits beyond which the disturbance is considered to be harmful to the load equipment

Table 8.2-2
Causes of Power Related Computer Failures, Northern Virginia, 1976[1]

Recorded Cause	Disturbance: Undervoltage	Disturbance: Outage	Number of Computer Failures
Wind and lightning	37	14	51
Utility equipment failure	8	0	8
Construction or traffic accident	8	2	10
Animals	5	1	6
Tree limbs	1	1	2
Unknown	21	2	23
Totals	80	20	100

Key developed the *recommended voltage tolerance envelope* shown in Figure 8.2-1. The design goals illustrated are recommendations to computer manufacturers for implementation in new equipment.

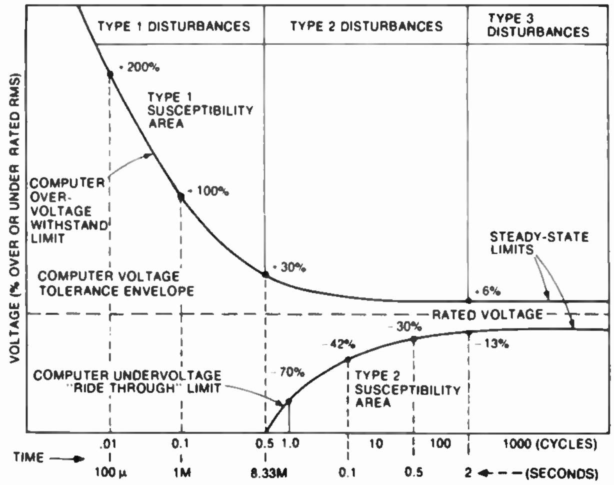


Figure 8.2-1. The recommended voltage tolerance envelope for computer equipment. This chart is based on pioneering work done by the Naval Facilities Engineering Command. The study identified how the magnitude *and* duration of a transient pulse must be considered in determining the damaging potential of a spike. The design goals illustrated in the chart are recommendations to computer manufacturers for implementation in new equipment. In the chart, cycles refer to cycles of 60 Hz ac.[1]

Assessing the Lightning Hazard

As identified by Key in his Naval Facilities study, the extent of lightning activity in an area significantly affects the probability of equipment failure caused by transient activity. The threat of a lightning flash to a facility is determined, in large part, by the type of installation and its geographic location. The type and character of the lightning flash are also important factors.

The *Keraunic number* of a geographic location describes the likelihood of lightning activity in that area. Figure 8.2-2 shows the *Isokeraunic map* of the United States, which estimates the number of lightning days per year across the country. On average, 30 storm days occur per year across the continental United States. This number does not fully describe the lightning threat because many individual lightning flashes occur during a single storm.

The structure of a facility has a significant effect on the exposure of equipment to potential lightning damage. Higher structures tend to collect and even trigger localized lightning flashes. Because storm clouds tend to travel at specific heights above the earth, conductive structures in mountainous areas more readily attract lightning activity. The *plant exposure factor* is a function of the size of the facility and the Isokeraunic rating of the area. The larger the physical size of an installation, the more likely it is to be hit by lightning during a storm. The longer a transmission line (ac or RF), the more lightning flashes it is likely to receive.

The relative frequency of power problems is seasonal in nature. As shown in Figure 8.2-3, most problems are noted during June, July and August. These high problem rates can be traced primarily to increased thunderstorm activity.

Table 8.2-3
Effects of Power System Configuration on Incidence of Computer Failures[1]

Configuration	Number of Disturbances: Undervoltage	Number of Disturbances: Outage	Recorded Failures
Overhead radial	12	6	18
Overhead spot network	22	1	23
Combined overhead (weighted[1])	16	4	20
Underground radial	6	4	10
Underground network	5	0	5
Combined underground (weighted[1])	5	2	7

[1] The combined averages weighted based on the length of time monitored (30 to 53 months).

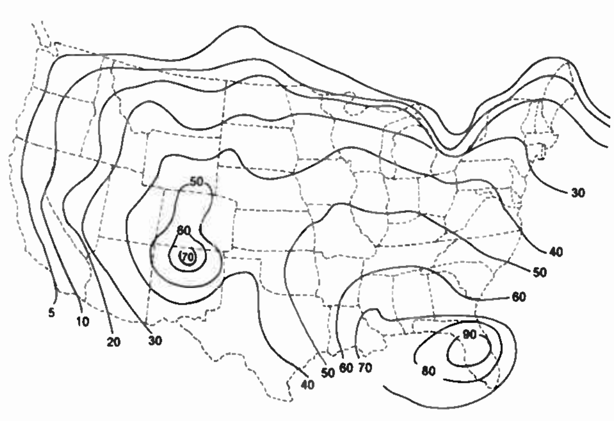


Figure 8.2-2. The Isokeraunic map of the United States, showing the approximate number of lightning days per year.

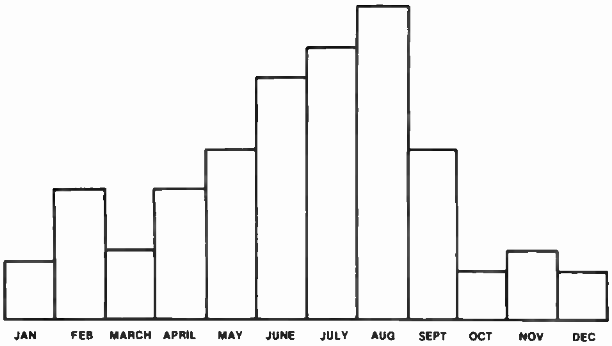


Figure 8.2-3. The relative frequency of power problems in the United States, classified by month.

FIPS Publication 94

In 1983, the U.S. Department of Commerce published a guideline summarizing the fundamentals of powering, grounding and protecting sensitive electronic devices. The document, known as *Federal Information Processing Standards Publication 94*, (FIPS Pub. 94) was first reviewed by governmental agencies and sent to the Computer Business Equipment Manufacturers Association (CBEMA) for review.[2] When the CBEMA group put its stamp of approval on the document, the data processing industry finally had an overarching guideline for power quality.

FIPS Pub. 94 was written to cover *automatic data processing equipment* (ADP), which at that time constituted the principal equipment that was experiencing difficulty running on normal utility supplied power. Since then, IEEE Standard P1100 was issued, which applies to all sensitive electronic equipment-including computer based broadcast systems. FIPS Pub. 94 is a guideline intended to provide a cost/benefit course of action. As a result, it can be relied upon to give the best solution to typical problems that will be encountered, for the least amount of money.

In addition to approving FIPS Pub. 94, the CBEMA group provided a curve that had been used as a guideline for their members in designing power supplies for modern electronic equipment. The CBEMA curve from the FIPS document is shown in Figure 8.2-4. (Note the similarity to the Key tolerance envelope shown in Figure 8.2-1.)

The curve is a susceptibility profile. In order better to explain its meaning, the curve has been simplified and redrawn in Figure 8.2-5. The vertical axis of the graph is the percent of voltage that is applied to the power circuit, and the horizontal axis is the time factor involved (in µs to s). In the center of the chart is the acceptable operating area, and on the outside of that area is a danger area on top and bottom. The danger zone at the top is a function of the tolerance of equipment to excessive voltage levels. The danger zone on the bottom sets the tolerance of equipment to a loss or reduction in applied power. The CBEMA guideline states that if the voltage supply stays within the accept-

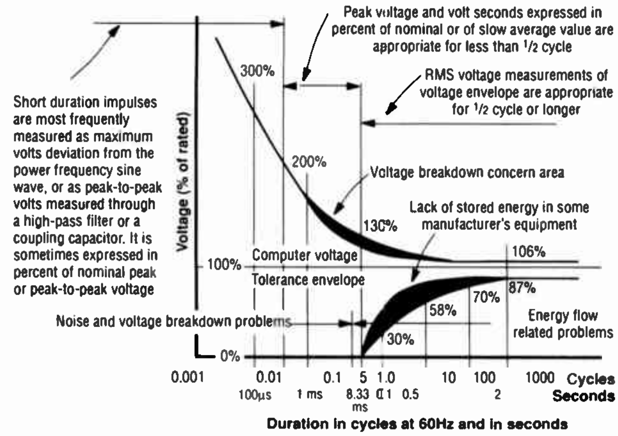


Figure 8.2-4. The CBEMA curve from FIPS Pub. 94.[2]

able area given by the curve, the sensitive load equipment will operate as intended.

Transient Protection Alternatives

A facility can be protected from transient disturbances in two basic ways: the *systems* approach or the *discrete device* approach. Table 8.2-4 outlines the major alternatives available:

- Uninterruptible power system (UPS) and standby generator
- UPS stand-alone system
- Secondary ac *spot network* where power is delivered to a facility via separate paths from the utility company
- Secondary *selective ac network* which is identical to the secondary ac spot network, except that a static transfer switch is included to permit load switching without interruption in service

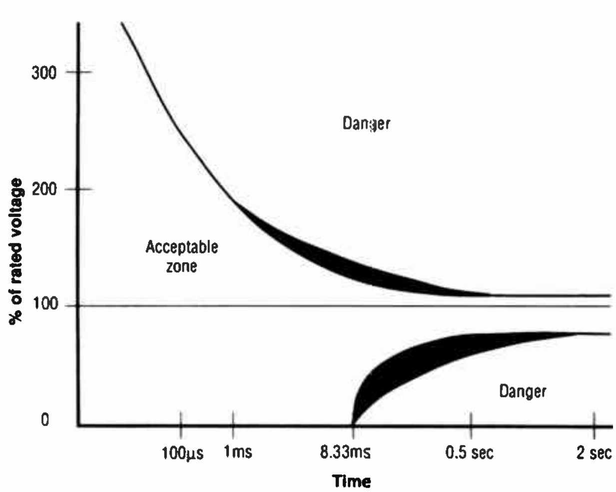


Figure 8.2-5. A simplified version of the CBEMA curve. Voltage levels outside the *acceptable zone* results in potential system shutdown and hardware and software loss.[2]

Table 8.2-4
Types of System-wide Protection Equipment Available to Facility Managers and the ac Line Abnormalities That Each Approach Can Handle[1,2]

System	Disturbance[3] Type 1	Disturbance Type 2	Disturbance Type 3
UPS system and standby generator	All source transients; no load transients	All	All
UPS system	All source transients; no load transients	All	All outages shorter than the battery supply discharge time
Secondary spot network[1]	None	None	Most, depending on the type of outage
Secondary selective network[2]	None	Most	Most, depending on the type of outage
Motor-generator set	All source transients; no load transients	Most	Only brown-out conditions
Shielded isolation transformer	Most source transients; no load transients	None	None
Suppressors, filters, lightning arrestors	Most transients	None	None
Solid state line voltage regulator/filter	Most source transients; no load transients	Some, depending on the response time of the system	Only brown-out conditions

[1] Dual power feeder network.
[2] Dual power feeder network using a static (solid state) transfer switch.
[3] Disturbance types illustrated graphically in Figure 8.2-1.

- Motor-generator set
- Shielded isolation transformer
- Suppressors, filters, and lightning arrestors
- Solid-state line voltage regulator/filter

Table 8.2-5 lists the relative benefits of each protection method. Because each installation is unique, a thorough investigation of facility needs should be conducted before purchasing any equipment. The systems approach offers the advantages of protection engineered to a particular application and need, and (usually) high level factory support during equipment design and installation. The systems approach also means higher costs for the end user.

Specifying System-Protection Hardware

Developing specifications for system-wide power conditioning/backup hardware requires careful analysis of various factors before a particular technology or a specific vendor is selected. Key factors in this process relate to the load hardware and load application. The electrical power required by a sensitive load may vary widely, depending on the configuration of the system. The principle factors that apply to system specification include the following:

- Power requirements, including: voltage, current, power factor, harmonic content and transformer configuration
- Voltage regulation requirements of the load
- Frequency stability required by the load, and the maximum permissible *slew rate* (the rate of change of frequency per second)
- Effects of unbalanced loading
- Overload and inrush current capacity
- Bypass capability
- Primary/standby path transfer time
- Maximum standby power reserve time
- System reliability and maintainability
- Operating efficiency.

An accurate definition of *critical applications* will aid in the specification process for a given site. The potential for future expansion also must be considered in all plans.

Power requirements can be determined either by measuring the actual installed hardware or by checking the nameplate ratings. Most nameplate ratings include significant safety margins. Moreover, the load normally will include a *diversity factor*; all individual elements of the load will not necessarily be operating at the same time.

Every load has a limited tolerance to noise and harmonic distortion. *Total harmonic distortion* (THD) is a measure of the quality of the waveform applied to the load. It is calculated by taking the geometric sum of the harmonic voltages present in the waveform, and expressing that value as a percentage of the fundamental voltage. Critical loads typically can withstand 5% THD, where no single harmonic exceeds 3%. The power conditioning system must provide this high quality output waveform to the load, regardless of the level of noise and/or distortion present at the ac input terminals.

If a power conditioning/standby system does not operate with high reliability, the results often can be disastrous. In addition to threats to health and safety, there is a danger of lost revenue and hardware damage. Reliability must be considered from three different viewpoints:

- Reliability of utility ac power in the area
- Impact of line voltage disturbances on computer based loads
- Ability of the protection system to maintain reliable operation when subjected to expected and unexpected external disturbances.

The environment in which the power conditioning system operates will have a significant effect on reliability. Extremes of temperature, altitude, humidity and vibration can be encountered in various applications. Extreme conditions can precipitate premature component failure and unexpected system shutdown. Most power protection equipment is rated for operation from 0°C to 40°C. During a commercial power failure, however, the ambient temperature of the equipment room

Table 8.2-5
Relative Merits of System-wide Protection Equipment[1,2]

System	Strong Points	Weak Points	Technical Profile
UPS system and standby generator	Full protection from power outage failures and transient disturbances; ideal for critical technical and life-safety loads	Hardware is expensive and may require special construction; electrically and mechanically complex; noise may be a problem; high annual maintenance costs	Efficiency 80-90%; typical high impedance presented to the load may be a consideration; frequency stability good; harmonic distortion determined by UPS system design
UPS system	Completely eliminates transient disturbances; eliminates surge and sag conditions; provides power outage protection up to the limits of the battery supply; ideal for critical load applications	Hardware is expensive; depending on battery supply requirements, special construction may be required; noise may be a problem; periodic maintenance required	Efficiency 80-90%; typical high impedance presented to the load may be a consideration; frequency stability good; harmonic content determined by inverter type
Secondary spot network[1]	Simple; inexpensive when available in a given area; protects against local power interruptions; no maintenance required by user	Not available in all locations; provides no protection from area-wide utility failures; provides no protection against transient disturbances or surge/sag conditions	Virtually no loss, 100% efficient; presents low impedance to the load; no effect on frequency or harmonic content
Secondary selective network[2]	Same as above; provides faster transfer from one utility line to the other	Same as above	Same as above
Motor-generator set	Electrically simple; reliable power source; provides up to 0.5 s power-fail ride-through in basic form; completely eliminates transient and surge/sag conditions	Mechanical system requires regular maintenance; noise may be a consideration; hardware is expensive; depending on m-g set design, power-fail ride-through may be less than typically quoted by manufacturer	Efficiency 80-90%; typical high impedance presented to the load may be a consideration; frequency stability may be a consideration, especially during momentary power-fail conditions; low harmonic content
Shielded isolation transformer	Electrically simple; provides protection against most types of transients and noise; moderate hardware cost; no maintenance required	Provides no protection from brown-out or outage conditions	No significant loss, essentially 100% efficient; presents low impedance to the load; no effect on frequency stability; usually low harmonic content
Suppressors, filters, lightning arrestors	Components inexpensive; units can be staged to provide transient protection exactly where needed in a plant; no periodic maintenance required	No protection from Type 2 or 3 disturbances; transient protection only as good as the installation job	No loss, 100% efficient; some units subject to power-follow conditions; no effect on impedance presented to the load; no effect on frequency or harmonic content
Solid state line voltage regulator/ filter	Moderate hardware cost; uses a combination of technologies to provide transient suppression and voltage regulation; no periodic maintenance required	No protection against power outage conditions; slow response time may be experienced with some designs	Efficiency 92-98%; most units present low impedance to the load; usually no effect on frequency; harmonic distortion content may be a consideration

[1] Dual power feeder network
[2] Dual power feeder network using a static (solid state) transfer switch

can easily exceed either value, depending on the exterior temperature. Operating temperature derating typically is required for altitudes in excess of 1000 ft.

Table 8.2-6 lists key power-quality attributes that should be considered when assessing the need for power-conditioning hardware for a broadcast facility.

MOTOR GENERATOR SET

As the name implies, a *motor-generator* (m-g) set consists of a motor powered by the ac utility supply that is mechanically tied to a generator, which feeds the load (see Figure 8.2-6). Transients on the utility line will have no effect on the load when this arrangement is used. Adding a flywheel to the motor-to-generator shaft will protect against brief power dips (up to 1/2 s on many models). Figure 8.2-7 shows the construction of a typical m-g set. The attributes of an m-g include the following:

- An independently generated source of voltage can be regulated without interaction with line voltage changes on the power source. Utility line changes of ±20% commonly can be held to within ±1% at the load
- The rotational speed and inertial momentum of the rotating mass represents a substantial amount of stored rotational energy, preventing sudden changes in voltage output when the input is momentarily interrupted
- The input and output windings are separated electrically, preventing transient disturbances from propagating from the utility company ac line to the load
- Stable electrical characteristics for the load include output voltage and frequency regulation, ideal sine wave output and true 120° phase shift for three-phase models
- Reduced problems relating to the power factor presented to the utility company power source.

Table 8.2-6
Power-Quality Attributes for Computer-Based Processing Hardware[2]

Environmental attribute	Typical environment	Acceptable limits for computer based systems: Normal	Acceptable limits for computer based systems: Critical
Line frequency	±0.1 to ±3%	±1%	±0.3%
Rate of frequency change	0.5 to 20 Hz/s	1.5 Hz/s	0.3 Hz/s
Over- and under-voltage	(± 5% to +6, −13.3%	+5% to −10%	± 3%
Phase imbalance	2 to 10%	5% max	3% max
Tolerance to low power factor	0.85 to 0.6 lagging	0.8 lagging	less than 0.6 lagging, or 0.9 leading
Tolerance to high steady state peak current	1.3 to 1.6 peak, rms	1.0 to 2.5 peak, rms	Greater than 2.5 peak, rms
Harmonic voltages	0 to 20% total rms	10 to 20% total, 5 to 10% largest	5% max total, 3% largest
Voltage deviation from sine wave	5 to 50%	5 to 10%	3 to 5%
Voltage modulation	Negligible to 10%	3% max	1% max
Surge/sag conditions	+10%, −15%	+20%, −30%	+5%, −5%
Transient impulses	2 to 3 times nominal peak value (0 to 130% Vs)	Varies; 1.0 to 1.5 kV typical	Varies; 200 to 500 V typical
RFI/EMI normal and common modes	10 V up to 20 kHz, less at high freq.	Varies widely, 3 V typical	Varies widely, 0.3 V typical
Ground currents	0 to 10 A plus impulse noise current	0.001 to 0.5 A or more	0.0035 A or less

The efficiency of a typical m-g ranges from 65–89%, depending on the size of the unit and the load. Motor-generator sets have been used widely to supply 415 Hz power to mainframe computers that require this frequency.

System Configuration

There are a number of types of m-g sets, each having its own characteristics, advantages and disadvantages. A simplified schematic diagram of an m-g is shown in Figure 8.2-8. The type of motor that drives the set is an important design element. Direct current motor drives can be controlled in speed independently of the frequency of the ac power source from which the dc is derived. Use of a dc motor, thereby, gives the m-g set the capability to produce power at the desired output frequency, regardless of variations in input frequency. The requirement for rectifier conversion hardware, control equipment and commutator maintenance are drawbacks to this approach that must be considered.

The simplest and least expensive approach to rotary power conditioning involves the use of an induction motor as the mechanical source. Unfortunately, the rotor of an induction motor turns slightly slower than the rotating field produced by the power source. This results in the generator being unable to produce 60 Hz output power if the motor is operated at 60 Hz and the machines are directly coupled end-to-end at their shafts, or are belted in a 1:1 ratio. Furthermore, the shaft speed and output frequency of the generator decreases as the load on the generator is increased. This potential for varying output frequency may be acceptable where the m-g set is used solely as the input to a power supply in which the ac is rectified and converted to dc. However, certain loads cannot tolerate frequency changes greater than 1 Hz/s and frequency deviations of more than 0.5 Hz from the nominal 60 Hz value.

Low slip induction motor driven generators are available that can produce 59.7 Hz at full load, assuming 60 Hz input. During power interruptions, the output frequency will drop further depending upon the length of the interruption. The capability of the induction motor to restart after a momentary power interruption is valuable. Various systems of variable speed belts have been tried successfully. Magnetically controlled

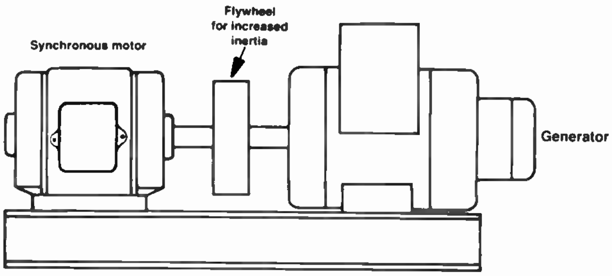


Figure 8.2-6. Two-machine motor-generator set with an optional flywheel to increase inertia and carry through capability.

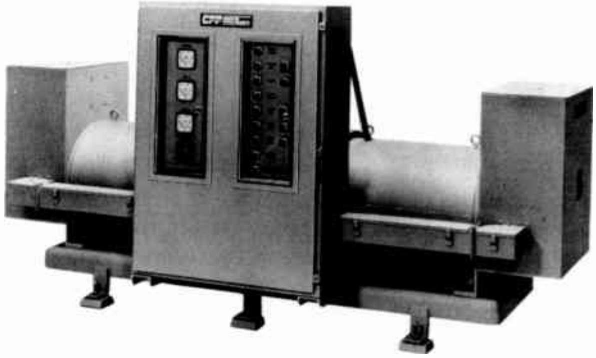


Figure 8.2-7. Construction of a typical m-g set. (*Courtesy of Computer Power Protection.*)

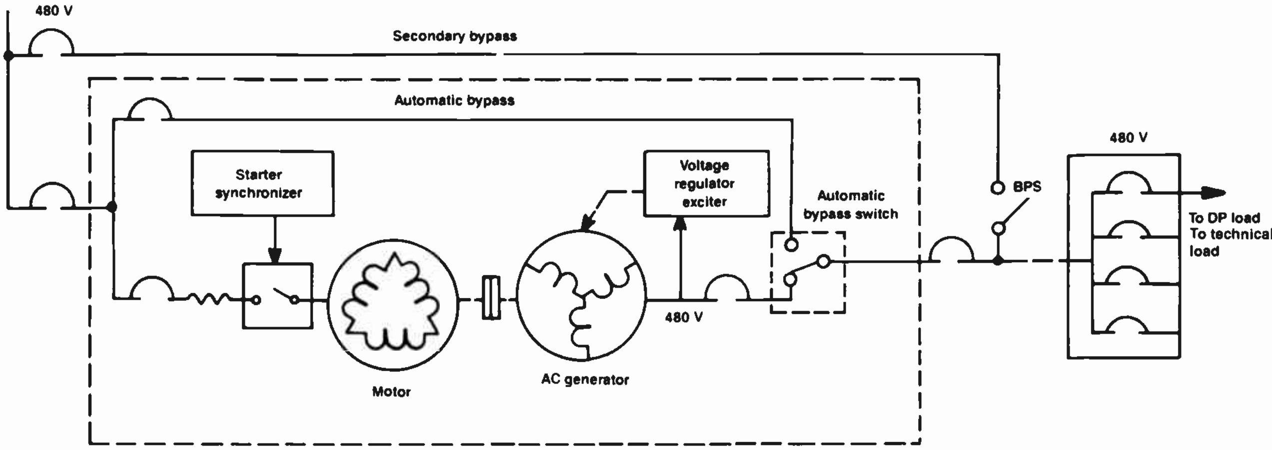


Figure 8.2-8. Simplified schematic diagram of an m-g set with automatic and secondary bypass capability.[2]

slipping clutches have been found to be largely unsatisfactory. Other approaches to make the induction motor drive the load at constant speed have produced mixed results.

Using a synchronous motor with direct coupling or a cogged 1:1 ratio belt drive guarantees that the output frequency will be equal to the motor input frequency. Although the synchronous motor is more expensive, it is more efficient and can be adjusted to provide a unity PF load to the ac source. The starting characteristics and the mechanical disturbance following a short line voltage interruption depends, to a large extent, on motor design. Many synchronous motors that are not required to start under load have weak starting torque, and may use a *pony motor* to aid in starting (see Figure 8.2-9). Those motors designed to start with a load have starting pole face windings that provide starting torque comparable to that of an induction motor. Such motors can be brought into synchronism while under load with proper selection of the motor and automatic starter system. Typical utility company ac interruptions are a minimum of six cycles (0.1 s). Depending upon the design and size of the flywheel used, the ride through period can be as much as 0.5 s or more. The generator will continue to produce output power for a longer

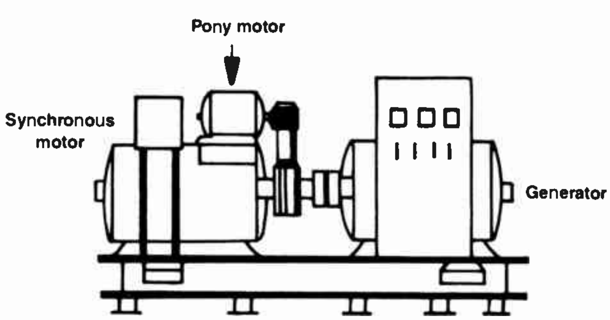


Figure 8.2-9. Use of a pony motor for an m-g set to aid in system starting and restarting.[2]

duration, but the frequency and rate of frequency change will most likely fall outside of the acceptable range of most technical loads after 0.5 s.

If input power is interrupted and does not return before the output voltage and frequency begin to fall outside acceptable limits, the generator output controller can be programmed to disconnect the load. Before this event, a warning signal is sent to the computer system control circuitry to warn of impending shutdown and to initiate an orderly interruption of active programs. This facilitates easy restart of the computers and associated network(s) after the power interruption has passed.

It is important for users to accurately estimate the length of time that the m-g set will continue to deliver acceptable power without input to the motor from the utility company. This data facilitates accurate power-fail shutdown routines. It is also important to ensure that the m-g system can handle the return of power without operating overcurrent protection devices because of high inrush currents that may be required to accelerate and synchronize the motor with the line frequency. Protection against the latter problem requires proper programming of the synchronous motor controller to correctly disconnect and then reconnect the field current supply.

Generators typically used in m-g sets have substantially higher internal impedance than equivalent kVA-rated transformers. Because of this situation, m-g sets sometimes are supplied with an oversize generator that will be lightly loaded, coupled with a smaller motor that is adequate to drive the actual load. This approach reduces the initial cost of the system, decreases losses in the motor, and provides a lower operating impedance for the load.

The most common utility supply voltage used to drive the input of an m-g set is 480 V. The generator output for systems rated at about 75 kVA or less is typically 208 Y/120 V. For larger loads, the most economical generator output is typically 480 V. A 480-208 Y/120 V three-phase isolating transformer usually is included to provide 208 Y/120 V power to the load equipment.

Single-Shaft Systems

There are two basic m-g set machine mechanical designs used for critical load applications: separate motor-generator systems and single-shaft, single-housing units. Both designs can use either a synchronous or induction motor. In each case, there are advantages and disadvantages. The separate machine design (discussed previously) uses a motor driving a physically separate generator by means of a coupling shaft or pulley. In an effort to improve efficiency and reduce costs, manufacturers also have produced various types of single-shaft systems.

The basic concept of a single-shaft system is to combine the motor and generator elements into a single unit. A common stator eliminates a number of individual components, making the machine less expensive to produce and mechanically more efficient. The common stator set substantially reduces mechanical energy losses associated with traditional m-g designs, and it improves system reliability as well. In one design, the stator is constructed so that alternate slots are wound with input and output windings. When it is fed with a three-phase supply, a rotating magnetic field is created, causing the dc-excited rotor to spin at a synchronous speed. By controlling the electrical characteristics of the rotor, control of the output at the secondary stator windings is accomplished.

Common stator machines offer lower working impedance for the load than a comparable two-machine system. For example, a typical 400 kVA machine has approximately an 800 kVA frame size. The larger frame size yields a relatively low impedance power source capable of clearing subcircuit fuses under fault conditions. The output of the unit typically can supply up to seven times the full load current under fault conditions. Despite the increase in frame size, the set is smaller and lighter than comparable systems because of the reduced number of mechanical parts.

Flywheel Considerations

In an effort to achieve higher energy and power densities, m-g set designers have devoted considerable attention to the flywheel element itself. New composite materials and power electronics technologies have resulted in compact flywheel batteries capable of high linear velocity at the outside radius of the flywheel (*tip speed*).[3] The rotational speed of the flywheel is important because the stored energy in a flywheel is proportional to the square of its rotational speed. Therefore, an obvious method for maximizing stored energy is to increase the speed of the flywheel. All practical designs, however, have a limiting speed, which is determined by the stresses developed within the wheel resulting from inertial loads. These loads are also proportional to the square of rotational speed. Flywheels built of composite materials weigh less and hence develop lower inertial loads at a given speed. In addition, composites are often stronger than conventional engineering metals, such as steel. This combination of high strength and low weight enables extremely high tip speeds, relative to conventional wheels.

For a given geometry, the limiting *energy density* (energy per unit mass) of a flywheel is proportional to the ratio of material strength to weight density, otherwise known as the *specific strength*. Table 8.2-7 illustrates the advantage that composite materials offer in this respect.

Recent advances in composite materials technology may allow nearly an order of magnitude advantage in the specific strength of composites when compared to even the best common engineering metals. The result of this continuous research in composites has been flywheels capable of operation at rotational speeds in excess of 100,000 rpm, with tip speeds in excess of 1,000 m/s.

These high speeds bring with them new challenges. The ultrahigh rotational speeds that are required to store significant kinetic energy in these systems virtually rule out the use of conventional mechanical bearings. Instead, most systems run on magnetic bearings. This relatively recent innovation uses magnetic forces to levitate a rotor, eliminating the frictional losses inherent in rolling element and fluid film bearings. Unfortunately, aerodynamic drag losses force most high speed flywheels to operate in a partial vacuum, which complicates the task of dissipating the heat generated by ohmic losses in the bearing electromagnets and rotor. In addition, active magnetic bearings are inherently unstable, and require sophisticated control systems to maintain proper levitation.

The integrated generator of these systems is usually a rotating field design, with the magnetic field supplied by rare earth permanent magnets. Because the specific strength of these magnets is typically just fractions of that of the composite flywheel, they must spin at much lower tip speeds; in other words, they must be placed very near the hub of the flywheel. This compromises the power density of the generator. An alternative is to mount them closer to the outer radius of the wheel, but contain their inertial loads with the composite wheel itself. Obviously, this forces the designer to either derate the machine speed or operate closer to the stress limits of the system.

Maintenance Considerations

Because m-g sets require some maintenance that necessitates shutdown, most systems provide bypass capability so the maintenance work can be performed

Table 8.2-7
Specific Strength of Selected Materials[4]

Material	Specific Strength (in^3)
Graphite/epoxy	3,509,000
Boron/epoxy	2,740,000
Titanium and alloys	1,043,000
Wrought stainless steel	982,000
Wrought high-strength steel	931,000
7000 series aluminum alloys	892,000

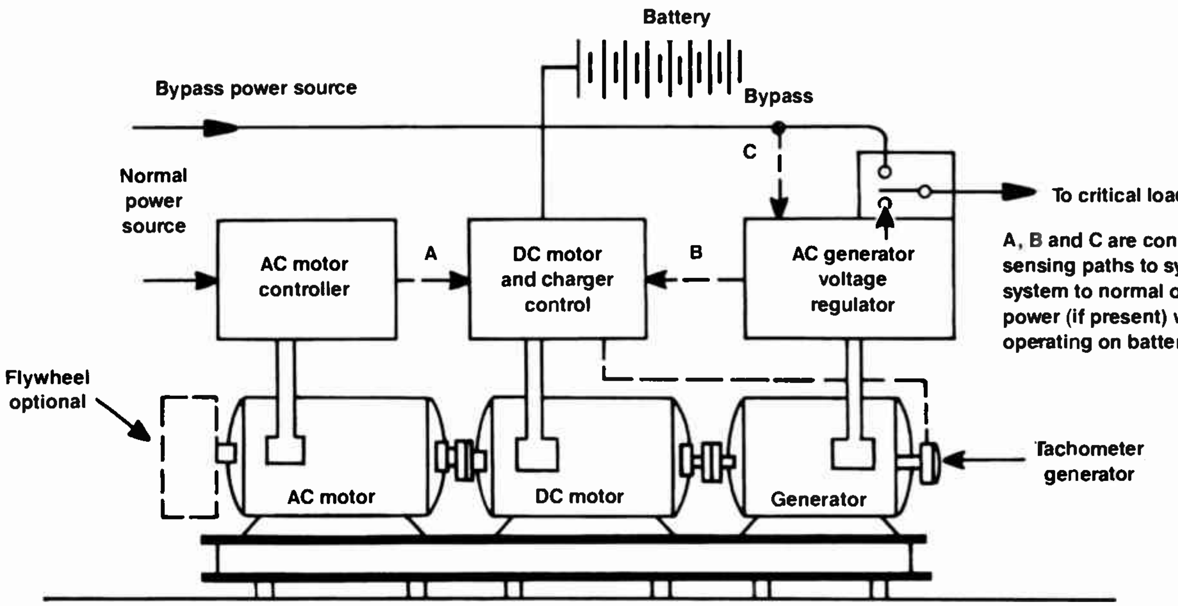


Figure 8.2-10. Uninterruptible m-g set with ac and dc motor drives.[2]

without having to take the load out of service. If the automatic bypass contactor, solid-state switch and control hardware are in the same cabinet as other devices that also need to be deenergized for maintenance, a secondary bypass is recommended. After the automatic bypass path has been established, transfer switching to the secondary bypass can be enabled, taking the m-g set and its automatic bypass system out of the circuit completely. Some automatic bypass control arrangements are designed to transfer the load of the generator to the bypass route with minimum disturbance. This requires the generator output to be synchronized with the bypass power before closing the switch and opening the generator output breaker. However, with the load taken off the generator, bypass power no longer will be synchronized with it. Consequently, retransfer of the load back to the generator may occur with some disturbance. Adjustment for minimum disturbance in either direction requires a compromise in phase settings, or a means to shift the phase before and after the transfer.

The use of rotating field exciters has eliminated the need for slip rings in most m-g designs. Brush inspection and replacement, therefore, are no longer needed. However, as with any rotating machinery, bearings must be inspected and periodically replaced.

Motor-Generator UPS

Critical applications that cannot tolerate even brief ac power interruptions can use the m-g set as the basis for an uninterruptible source of power through the addition of a battery-backed dc motor to the line-driven ac motor shaft. This concept is illustrated in Figure 8.2-10. The ac motor normally supplies power to drive the system from the utility company line. The shafts of the three devices all are interconnected, as shown in the Figure. When ac power is present, the dc motor serves as a generator to charge the battery bank. When line voltage is interrupted, the dc motor is powered by the batteries. Figure 8.2-11 shows a modified version of this basic m-g UPS using only a dc motor as the

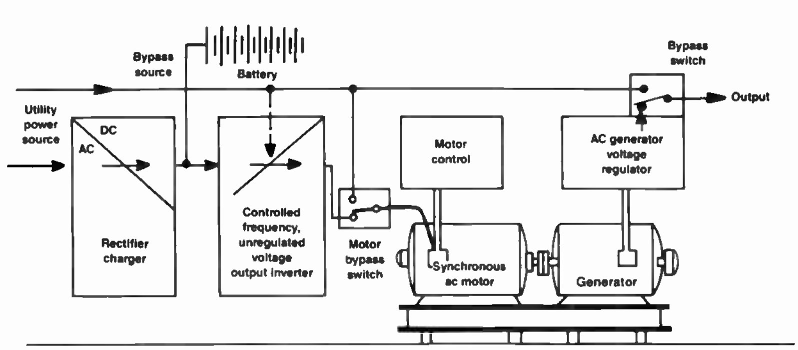


Figure 8.2-11. Uninterruptible m-g set using a single dc drive motor.[2]

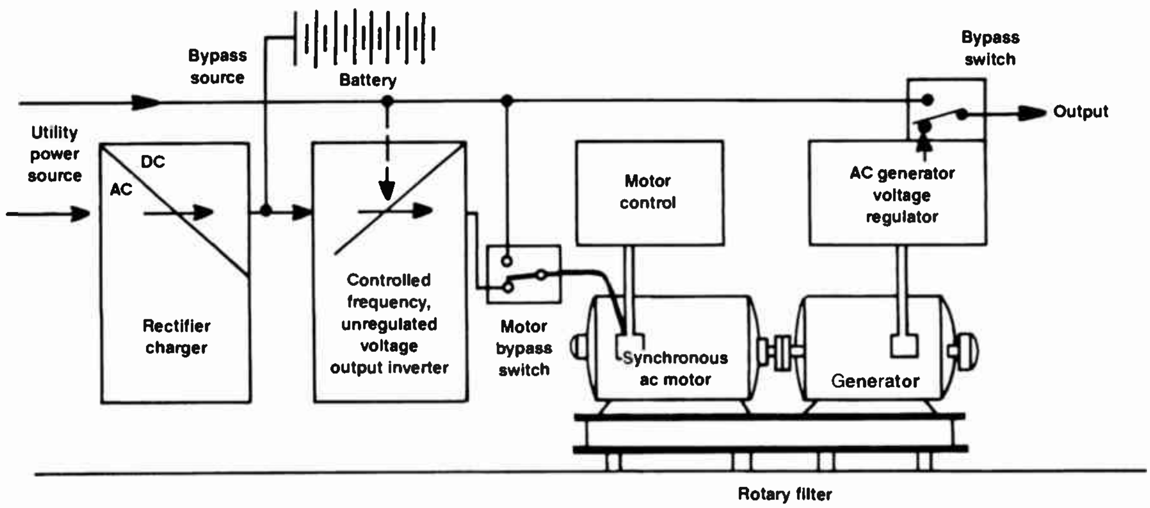


Figure 8.2-12. Uninterruptible m-g set using a synchronous ac motor.[2]

mechanical power source. This configuration eliminates the inefficiency involved in having two motors in the system. Power from the utility source is rectified to provide energy for the dc motor, plus power for charging the batteries. A complex control system to switch the ac motor off and the dc motor on in the event of a utility power failure is not needed in this design.

The m-g UPS also can be built around a synchronous ac motor, as illustrated in Figure 8.2-12. Utility ac energy is rectified and used to drive an inverter, which provides a regulated frequency source to power the synchronous motor. The output from the dc-to-ac inverter need not be a well formed sine wave, nor a well regulated source. The output from the generator will provide a well regulated sine wave for the load. The m-g set also can be operated in a bypass mode that eliminates the rectifier, batteries and inverter from the current path, operating the synchronous motor directly from the ac line.

Kinetic Battery Storage System

As outlined previously, one of the parameters that limits the ride through period of an m-g set is the speed decay of the flywheel/generator combination. As the flywheel slows down, the output frequency drops. This limits the useful ride through period to 0.5 s or so. Figure 8.2-13 shows an ingenious modification of the classic power conditioning template that extends the potential ride-through considerably. As shown in the Figure, an m-g set is used in a UPS based system as an element of the dc supply grid. The major components of the system include:

- Steel flywheel for energy storage
- Small drive motor, sized at 15-20% of the rated system output, to start the flywheel and maintain its normal operating speed
- Variable speed drive (VSD) to slowly ramp the flywheel up to speed and maintain it at the desired rpm
- Generator to convert the kinetic energy stored in the flywheel into electrical energy
- Diode bridge rectifier to convert the ac generator output to dc for use by the UPS bus, which continues to draw usable energy essentially independent of flywheel rpm.

Because the ac output of the generator is converted to dc, frequency is no longer a limiting factor, which allows the dc voltage output to be maintained over a much greater range of the flywheel's rpm envelope.

In operation, the small drive motor spins the flywheel while the variable speed drive maintains the proper motor speed.[5] Because the amount of stored kinetic energy increases by the square of the flywheel rpm, it is possible to greatly increase stored energy, and thus ride through, by widening the flywheel's usable energy output range. These factors permit typical ride through times from 10 s to several minutes, depending on loading and other operating conditions. Among the

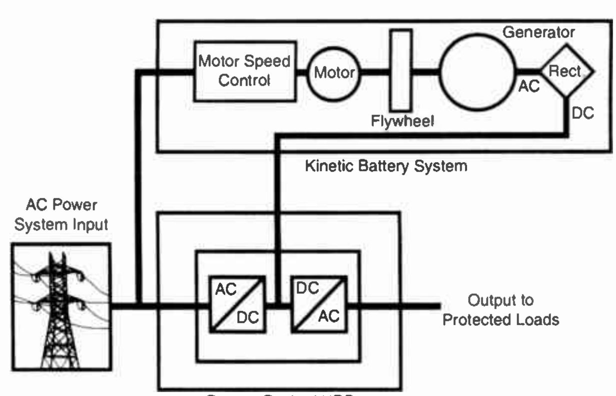


Figure 8.2-13. Functional block diagram of a kinetic battery m-g system for powering a UPS.[5]

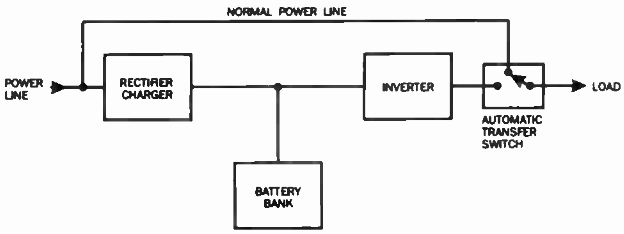


Figure 8.2-14. Forward-transfer UPS system.

benefits of this approach include the reduced cycling of battery supplies and engine-generators systems.

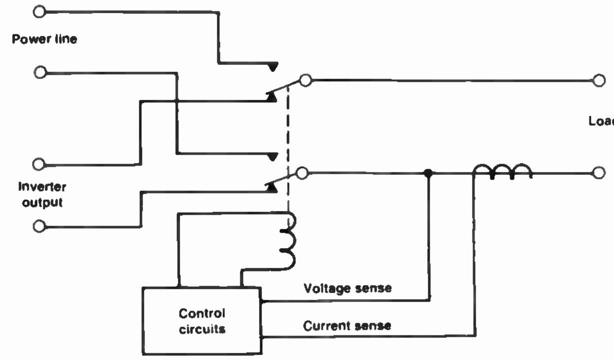


Figure 8.2-16. Electromechanical load-transfer switch.

UNINTERRUPTIBLE POWER SYSTEMS

Uninterruptible power systems (UPS) have become a virtual necessity for powering large or small computer systems where the application serves a critical need, such as broadcasting. Computers and data communications systems are no more reliable than the power from which they operate. The difference between UPS and *emergency standby power* is that the UPS is always in operation. It reverts to an alternative power source, such as the utility company, only if the UPS fails or needs to be deactivated for maintenance. Even then, the transfer of power occurs so quickly (within milliseconds) that it does not interrupt proper operation of the load.

Emergency standby power is normally off and does not start (manually or automatically) until the utility ac feed fails. A diesel generator can be started within 10 to 30 s if the system has been maintained properly. Such an interruption, however, is far too long for computer based hardware. Most systems cannot ride through more than 8 to 22 ms of power interruption. Systems that can successfully ride through short duration power breaks, as far as energy continuity is concerned, still may enter a fault condition because of electrical noise created by the disturbance.

UPS hardware is available in a number of different configurations. All systems, however, are variations of two basic designs:

- *Forward-transfer mode:* the load normally is powered by the utility power line and the inverter is idle. If a commercial power failure occurs, the inverter is started and the load is switched. This configuration is illustrated in Figure 8.2-14. The primary drawback of this approach is the lack of load protection from power line disturbances during normal (utility-powered) operation.
- *Reverse-transfer mode:* The load normally is powered by the inverter. In the event of an inverter failure, the load is switched directly to the utility line. This configuration is illustrated in Figure 8.2-15. The reverse-transfer mode is, by far, the most popular type of UPS system in use for large-scale systems.

The type of load-transfer switch used in the UPS system is another critical design parameter. The continuity of ac power service to the load is determined by the type of switching circuit used. An electromechanical transfer switch, shown in Figure 8.2-16, is limited to switch times of 20 to 50 ms. This time delay may cause sensitive load equipment to malfunction, and perhaps shut down. A control circuit actuates the relay when the sensed output voltage falls below a preset value, such as 94% of nominal. A static transfer switch, shown in Figure 8.2-17, can sense a failure and switch the load in about 4 ms. Most loads will ride through this short delay without any malfunction. To accomplish a smooth transition, the inverter output must be synchronized with the power line.

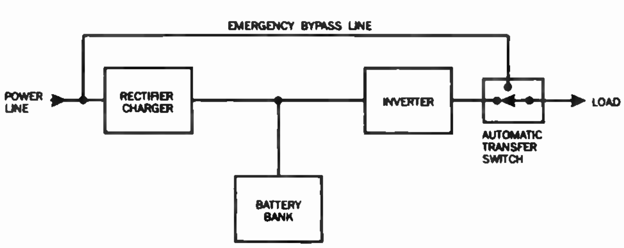


Figure 8.2-15. Reverse-transfer UPS system.

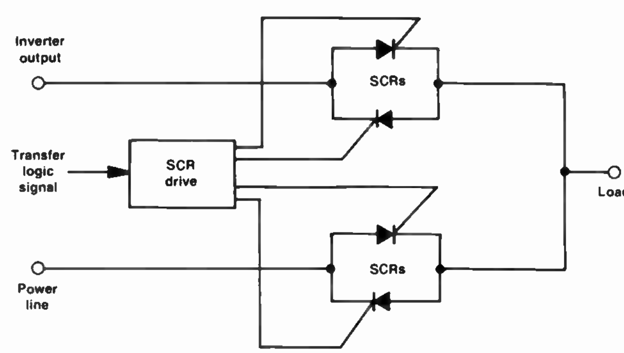


Figure 8.2-17. Static load-transfer switch.

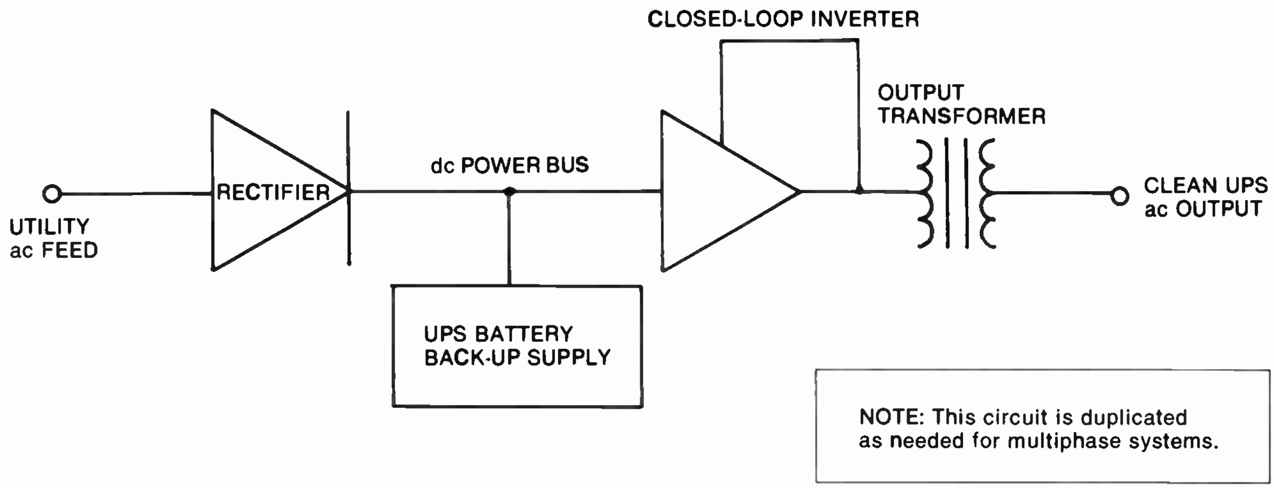


Figure 8.2-18. Block diagram of an uninterruptible power system using ac rectification to float the battery supply. A closed-loop inverter draws on this supply and delivers clean ac power to the protected load.

UPS Configuration

The basic uninterruptible power system is built around a battery-driven inverter, with the batteries recharged by the utility ac line. As shown in Figure 8.2-18, ac from the utility feed is rectified and applied to recharge or *float* a bank of batteries. This dc power drives a single- or multiphase closed-loop inverter, which regulates output voltage and frequency. The output of the inverter is generally a sine wave, or pseudo sine wave (a stepped square wave). If the utility voltage drops or disappears, current is drawn from the batteries. When ac power is restored, the batteries are recharged. Many UPS systems incorporate a standby diesel generator that starts as soon as the utility company feed is interrupted. With this arrangement, the batteries are called upon to supply operating current for only 30 s or so, until the generator gets up to speed. A UPS system intended to power a computer center, is illustrated in Figure 8.2-19.

Power-Conversion Methods

Solid-state UPS systems that do not employ rotating machinery utilize one of several basic concepts. The design of an inverter is determined primarily by the operating power level. The most common circuit configurations include:

- Ferroresonant inverter
- Delta magnetic inverter

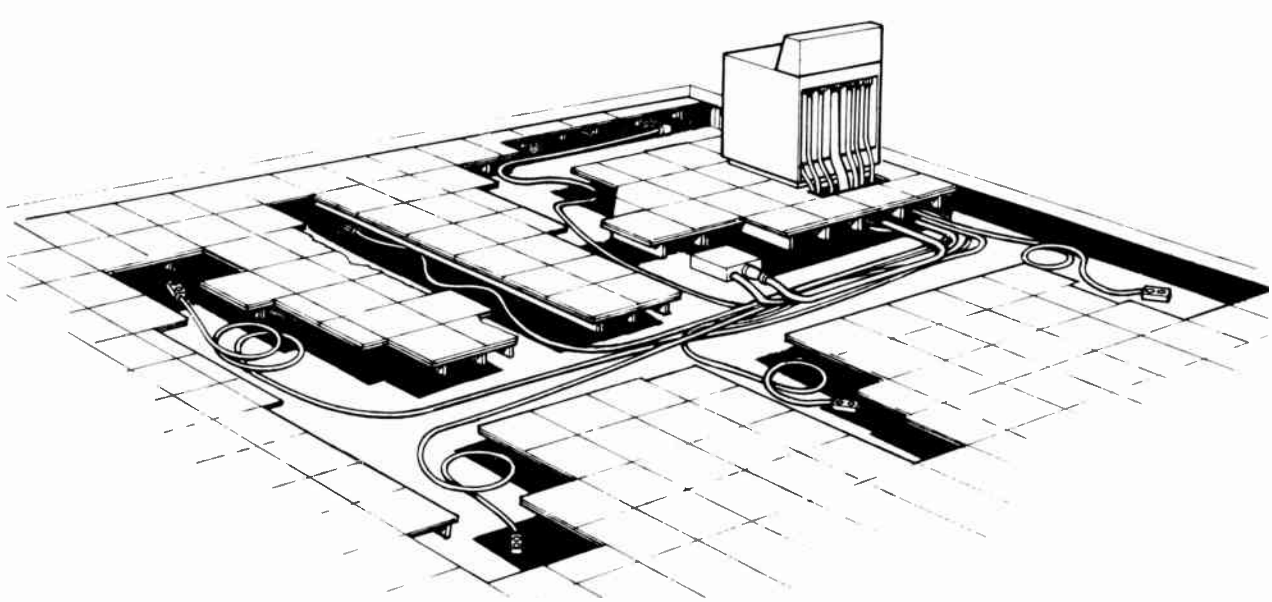


Figure 8.2-19. Installation details for a computer room UPS power conditioner.

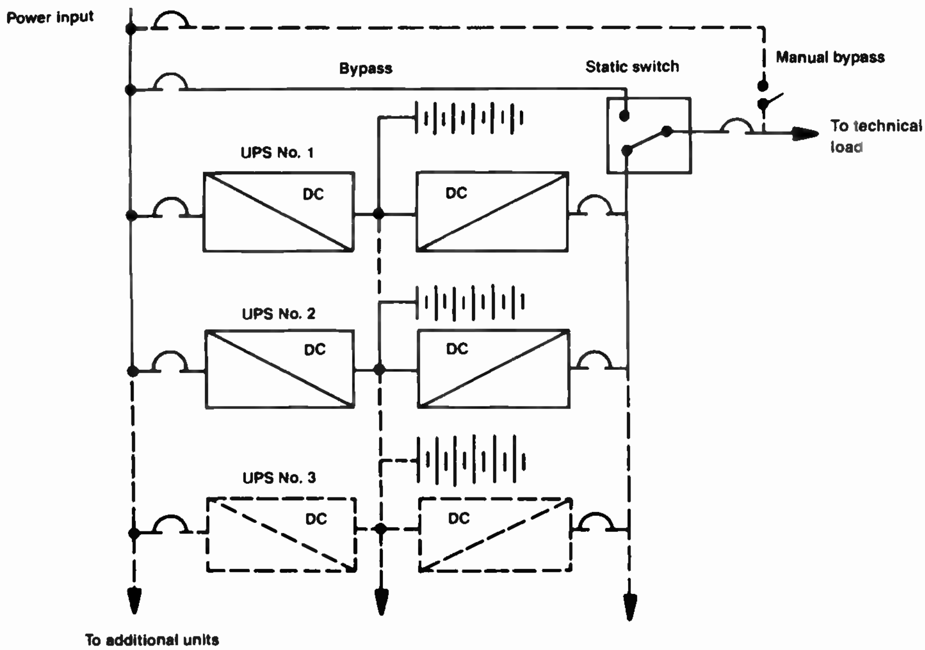


Figure 8.2-20. Configuration of a parallel redundant UPS system.[2]

- Inverter-fed *L/C* tank
- Quasi-square wave inverter
- Step wave inverter
- Pulse-width modulation inverter
- Phase modulation inverter

Redundant Operation

UPS systems can be configured as either a single, large power conditioning/backup unit or as several smaller systems arranged in a *parallel redundant* or *isolated redundant* mode. In the parallel redundant mode, the UPS outputs are connected together and share the total load (see Figure 8.2-20). The power-output ratings of the individual UPS systems are selected to provide for operation of the entire load with any one UPS unit out of commission. In the event of expansion of the broadcast facility, additional UPS units can be added to carry the load. The parallel system provides the ability to cope with the failure of any single unit.

An isolated redundant system, illustrated in Figure 8.2-21, divides the load among several UPS units. If one of the active systems fails, a static bypass switch will connect the affected load to a standby UPS system dedicated to that purpose. The isolated redundant system does not permit the unused capacity of one UPS unit to be utilized on a computer based system that is loading another UPS to full capacity. The benefit of the isolated configuration is its immunity to system wide failures.

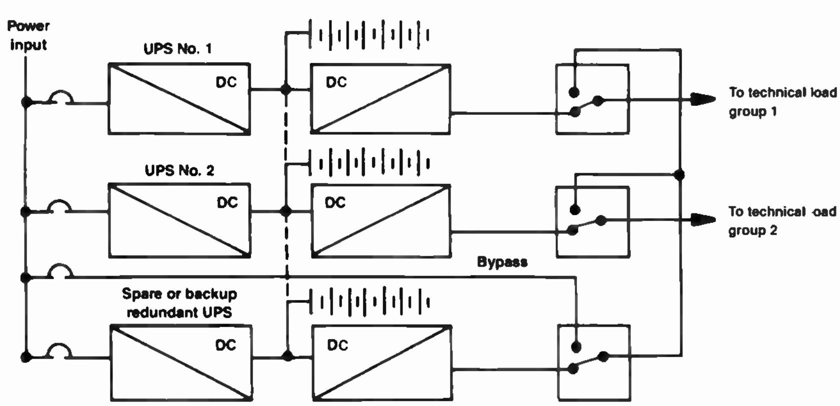


Figure 8.2-21. Configuration of an isolated redundant UPS system.[2]

Output Transfer Switch

Fault conditions, maintenance operations and system reconfiguration require the load to be switched from one power source to another. This work is accomplished with an output transfer switch. As discussed previously, most UPS systems use electronic (static) switching. Electromechanical or motor-driven relays operate too slowly for most loads. Static transfer switches can be configured as either of the following:

- *Break-before-make:* Power output is interrupted before transfer is made to the new source.
- *Make-before-break:* The two power sources are overlapped briefly so as to prevent any interruption in ac power to the load.

Figure 8.2-22 illustrates each approach to load switching.

For critical load applications, a make-before-break transfer is necessary. For the switchover to be accomplished with minimum disturbance to the load, both power sources must be synchronized. The UPS system must, therefore, be capable of synchronizing to the utility ac power line (or other appropriate power source).

Battery Supply

UPS systems typically are supplied with sufficient battery capacity to carry a technical load for periods ranging from 5 minutes to 1 hour or more. Long backup time periods usually are handled by a standby diesel generator. Batteries require special precautions. For large installations, they almost always are placed in a room dedicated to that purpose. Proper temperature control is important for long life and maximum discharge capacity.

Most rectifier/charger circuits operate in a *constant-current* mode during the initial charge period, and automatically switch to a *constant-voltage* mode near the end of the charge cycle. This provides maximum battery life consistent with rapid recharge. It also prevents excessive battery outgassing and water consumption. The charger provides a *float voltage level* for maintaining the normal battery charge, and sometimes a higher voltage to equalize certain devices.

Four battery types typically are found in UPS systems:

- **Semisealed lead calcium.** A gel-type electrolyte is used that does not require the addition of water. There is no outgassing or corrosion. This type of battery is used when the devices are integral to small UPS units, or when the batteries must be placed in occupied areas. The lifespan of a semisealed lead calcium battery, under ideal conditions, is about 5 years

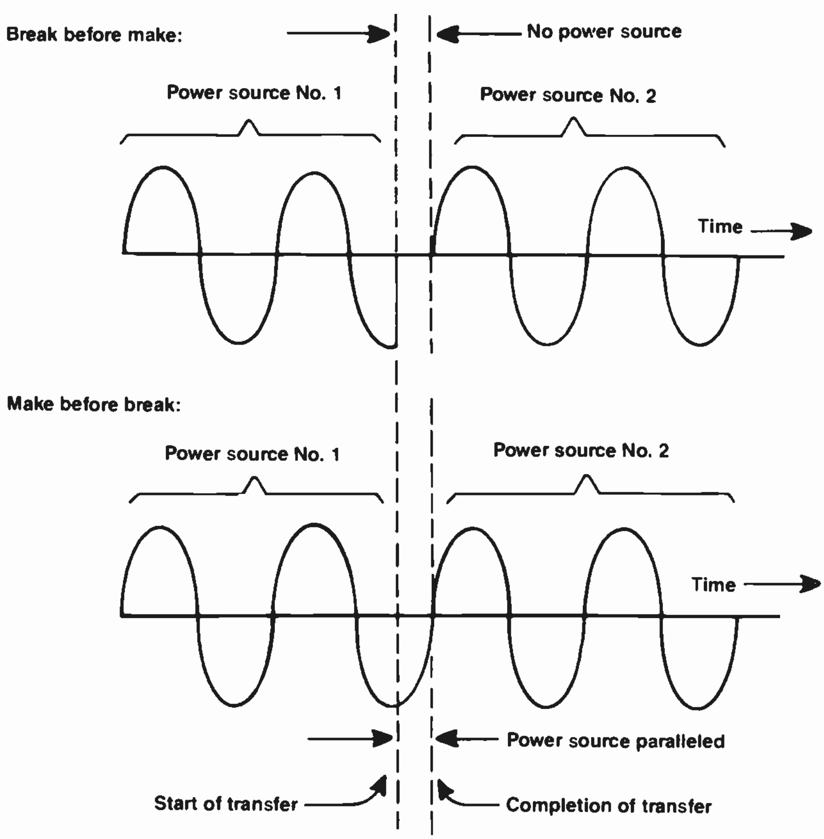


Figure 8.2-22. Static transfer switch modes.

- **Conventional lead calcium.** The most common battery type for UPS installations, these units require watering and terminal cleaning about every 6 months. Expected lifetime ranges up to 20 years. Conventional lead-calcium batteries outgas hydrogen under charge conditions and must be located in a secure, ventilated area
- **Lead-antimony.** The traditional lead-acid batteries, these devices are equivalent in performance to lead-calcium batteries. Maintenance is required every 3 months. Expected lifetime is about 10 years. To retain their capacity, lead-antimony batteries require a monthly equalizing charge
- **Nickel-cadmium.** Advantages of the nickel-cadmium battery includes small size and low weight for a given capacity. These devices offer excellent high and low temperature properties. Life expectancy is nearly that of a conventional lead-calcium battery. Nickel-cadmium batteries require a monthly equalizing charge, as well as periodic discharge cycles to retain their capacity. Nickel-cadmium batteries are the most expensive of the devices typically used for UPS applications.

DEDICATED PROTECTION SYSTEMS

A wide variety of power protection technologies are available to solve specific problems at a facility. The method chosen depends upon a number of factors, not the least of which is cost. Although UPS units and m-g sets provide a high level of protection against ac line disturbances, the costs of such systems are high. Many times, adequate protection can be provided using other technologies at a fraction of the cost of a full-featured, facility-wide system. The applicable technologies include:

- Ferroresonant transformer
- Isolation transformer
- Tap-changing regulator
- Line conditioner

Line Conditioner

A line conditioner combines the functions of an isolation transformer and a voltage regulator in one unit. The three basic types of line conditioners for sensitive load applications are:

- **Linear amplifier correction system.** As illustrated in Figure 8.2-23, correction circuitry compares the ac power output to a reference source, derived from a 60 Hz sine wave generator. A correction voltage is developed and applied to the secondary power winding to cancel noise and voltage fluctuations. A box shield around the primary winding provides common-mode impulse rejection (80 to 100 dB typical). The linear amplifier correction system is effective and fast, but the overall regulating range is limited
- **Hybrid ferroresonant transformer.** As shown in Figure 8.2-24, this system consists of a ferroresonant transformer constructed using circuit isolation techniques. The box and Faraday shields around the primary and compensating windings give the transformer essentially the characteristics of a noise-attenuating device, while preserving the volt-

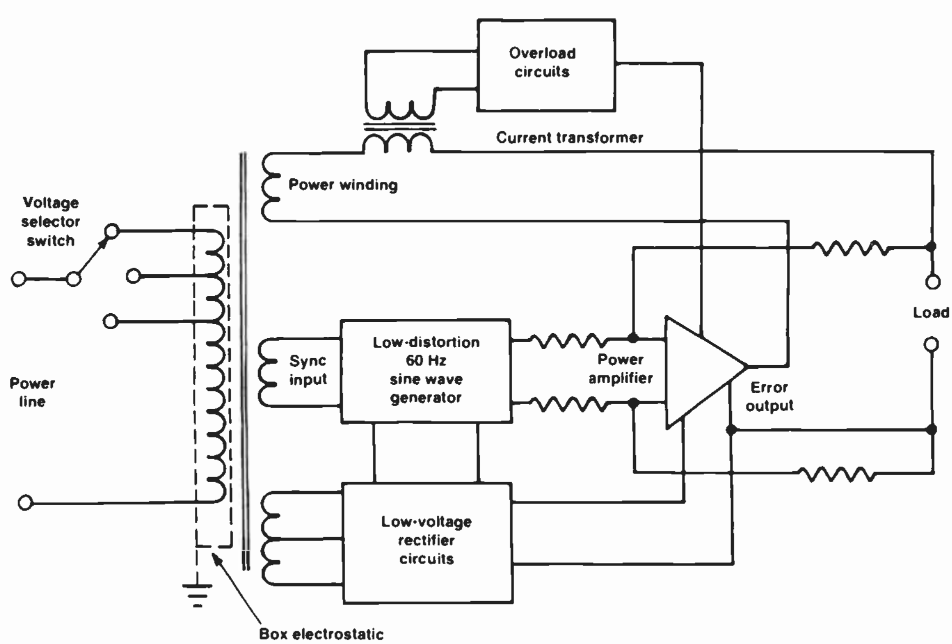


Figure 8.2-23. Power line conditioner using linear amplifier correction.

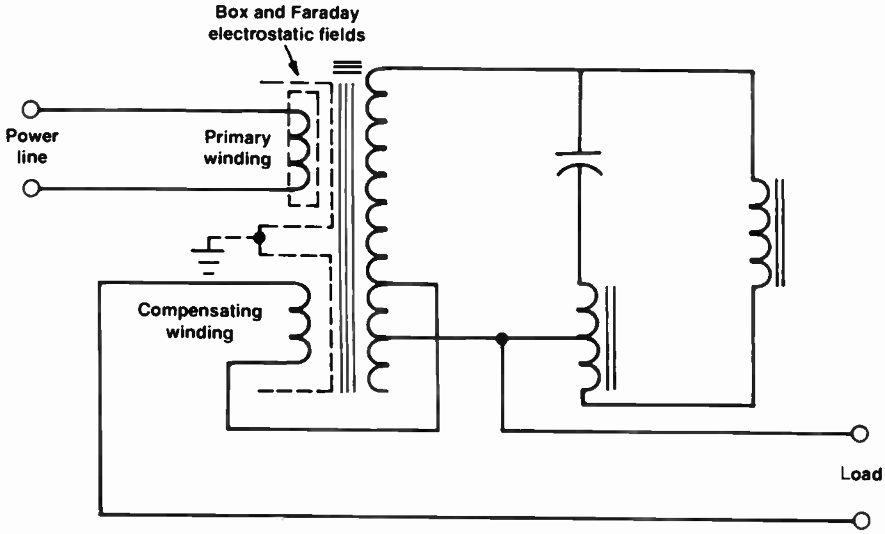


Figure 8.2-24. Line conditioner built around a shielded-winding ferroresonant transformer.

age regulating characteristics of a ferroresonant transformer

- **Electronic tap changing high isolation transformer.** This system is built around a high attenuation isolation transformer with a number of primary winding taps. SCR pairs control voltage input to each tap, as in a normal tap changing regulator. Tap changing also can be applied to the secondary, as shown in Figure 8.2-25. The electronic tap changing high isolation transformer is an efficient design that effectively regulates voltage output and prevents noise propagation to the load.

CONCLUSION

There are, thus, a number of approaches to ac power conditioning that are practical and applicable to broadcast facilities. This chapter only scratched the surface of this topic. Interested readers are referred to the *AC Power Systems Handbook*, 2nd Edition for a detailed examination of the subject.

In the realm of power quality, you get what you pay for. It is rarely practical to shield a facility from every conceivable disturbance that can occur on the ac input line. You can, however, fashion a practical

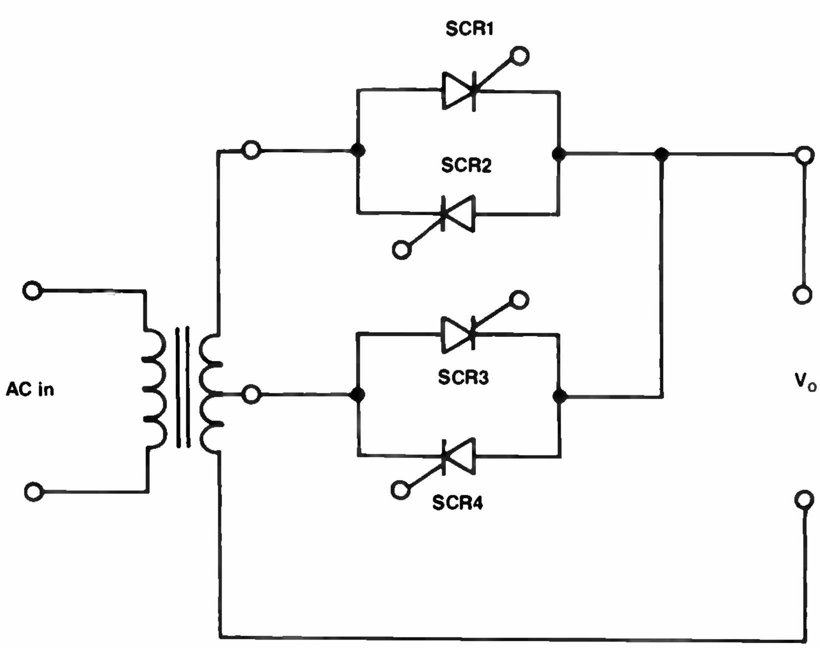


Figure 8.2-25. Secondary-side synchronous tap changing transformer.

and cost-effective protection system that meets the needs of the facility for maximum up-time. ′

REFERENCES

1. Key, Lt. Thomas: "The Effects of Power Disturbances on Computer Operation," IEEE Industrial and Commercial Power Systems Conference, Cincinnati, June 7, 1978.
2. Federal Information Processing Standards Publication No. 94, *Guideline on Electrical Power for ADP Installations*, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1983.
3. Plater, Bryan B., and James A. Andrews, "Advances in Flywheel Energy Storage Systems," *Power Quality "97*, Intertec International, Ventura, Calif., pp. 460–469, 1997.
4. "Materials Selector 1987, *Materials Engineering*, Penton Publishing.
5. Weaver, Eugene, J., "Dynamic Energy Storage System," *Power Quality Solutions/Alternative Energy*, Intertec International, Ventura, Calif., pp. 373–380, 1996.

BIBLIOGRAPHY

"How to Correct Power Line Disturbances," Dranetz Technologies, Edison, NJ, 1985.

Lawrie, Robert, *Electrical Systems for Computer Installations*, McGraw-Hill, New York, 1988.

Martzloff, F. D., "The Development of a Guide on Surge Voltages in Low-Voltage AC Power Circuits," 14th Electrical/Electronics Insulation Conference, IEEE, Boston, October 1979.

Newman, Paul, "UPS Monitoring: Key to an Orderly Shutdown," *Microservice Management*, Intertec Publishing, Overland Park, KS, March 1990.

Noise Suppression Reference Manual, Topaz Electronics, San Diego, CA.

Nowak, Stewart, "Selecting a UPS," *Broadcast Engineering*, Intertec Publishing, Overland Park, KS, April 1990.

Pettinger, Wesley, "The Procedure of Power Conditioning," *Microservice Management*, Intertec Publishing, Overland Park, KS, March 1990.

Smeltzer, Dennis, "Getting Organized About Power," *Microservice Management*, Intertec Publishing, Overland Park, KS, March 1990.

Whitaker, Jerry C., *AC Power Systems Handbook*, 2nd edition, CRC Press, Boca Raton, FL, 1998.

8.3
FACILITY GROUNDING PRACTICES

W.E. DEWITT
PURDUE UNIVERSITY, WEST LAFAYETTE, IN

INTRODUCTION

A facility can be defined as something that is built, installed or established to serve a particular purpose.[1] A facility is usually thought of as a single building or group of buildings. The National Electrical Code (NEC) uses the term *premises* to refer to a facility when it defines premises wiring as the interior and exterior (facility) wiring, such as power, lighting, control and signal systems. Premises wiring includes the service and all permanent and temporary wiring between the service and the load equipment. Premises wiring does not include wiring internal to any load equipment.

This chapter will define grounding and list the reasons for grounding; discuss equipment grounding, system grounding and earth connections; apply recommended facility grounding practice to the safe design and construction of electronics installations; and present grounding terminology. Applicable codes and industry standards will be emphasized.

Reasons for Grounding

The Institute of Electrical and Electronic Engineers (IEEE) defines grounding as a conducting connection, whether intentional or accidental, by which an electric circuit or equipment is connected to the earth, or to some conducting body of relatively large extent that serves in place of the earth. It is used for establishing and maintaining the potential of the earth (or of the conducting body) or approximately that potential, on conductors connected to it, and for conducting ground current to and from the earth (or the conducting body).[2]

Based on the IEEE definition, the reasons for grounding can be identified as:

- Personnel safety by limiting potentials between all noncurrent-carrying metal parts of an electrical distribution system all non-current-carrying metal parts of an electrical distribution system and the earth
- Personnel safety and control of electrostatic discharge (ESD) by limiting potentials between all noncurrent-carrying metal parts of an electrical distribution system and earth
- Fault isolation and equipment safety by providing a low-impedance fault return path to the power source to facilitate the operation of overcurrent devices during a ground fault.

The IEEE definition makes an important distinction between *ground* and *earth*. *Earth* refers to planet Earth and *ground* refers to the equipment grounding system, which includes equipment grounding conductors, metallic raceways, cable armor, enclosures, cabinets, frames, building steel and all other noncurrent carrying metal parts of the electrical distribution system.

There are other reasons for grounding not implicit in the IEEE definition. Overvoltage control has long been a benefit of power system grounding, and is described in IEEE Standard 142, *The Green Book*.[3] With the increasing use of electronic computer systems, noise control has become associated with the subject of grounding, and is described in IEEE Standard 1100, *The Emerald Book*.[4]

EQUIPMENT GROUNDING

Personnel safety is achieved by interconnecting all noncurrent-carrying metal parts of an electrical distribution system, and then connecting the interconnected metal parts to the earth. This process of interconnecting metal parts is called *equipment grounding* and is illustrated in Figure 8.3-1, where the equipment grounding conductor is used to interconnect the metal enclosures. Equipment grounding ensures that there is no difference of potential, and thus no shock hazard between noncurrent-carrying metal parts anywhere in the electrical distribution system. Connecting the equipment grounding system to earth ensures that there is no difference of potential between the earth and the equipment grounding system. It also prevents static charge buildup.

System Grounding

System grounding, which is also illustrated in Figure 8.3-1, is the process of intentionally connecting one

[1] Webster's New Collegiate Dictionary.

[2] IEEE Standard 100, *Definitions of Electrical and Electronic Terms*, The Institute of Electrical and Electronics Engineers, Inc., New York, NY.

[3] IEEE Standard 142, *Recommended Practice for Grounding Industrial and Commercial Power Systems*, The Institute of Electrical and Electronics Engineers, Inc., New York, NY, 1982.

[4] IEEE Standard 1100, *Recommended Practice for Powering and Grounding Sensitive Electronics Equipment*, The Institute of Electrical and Electronics Engineers, Inc., New York, NY, 1992.

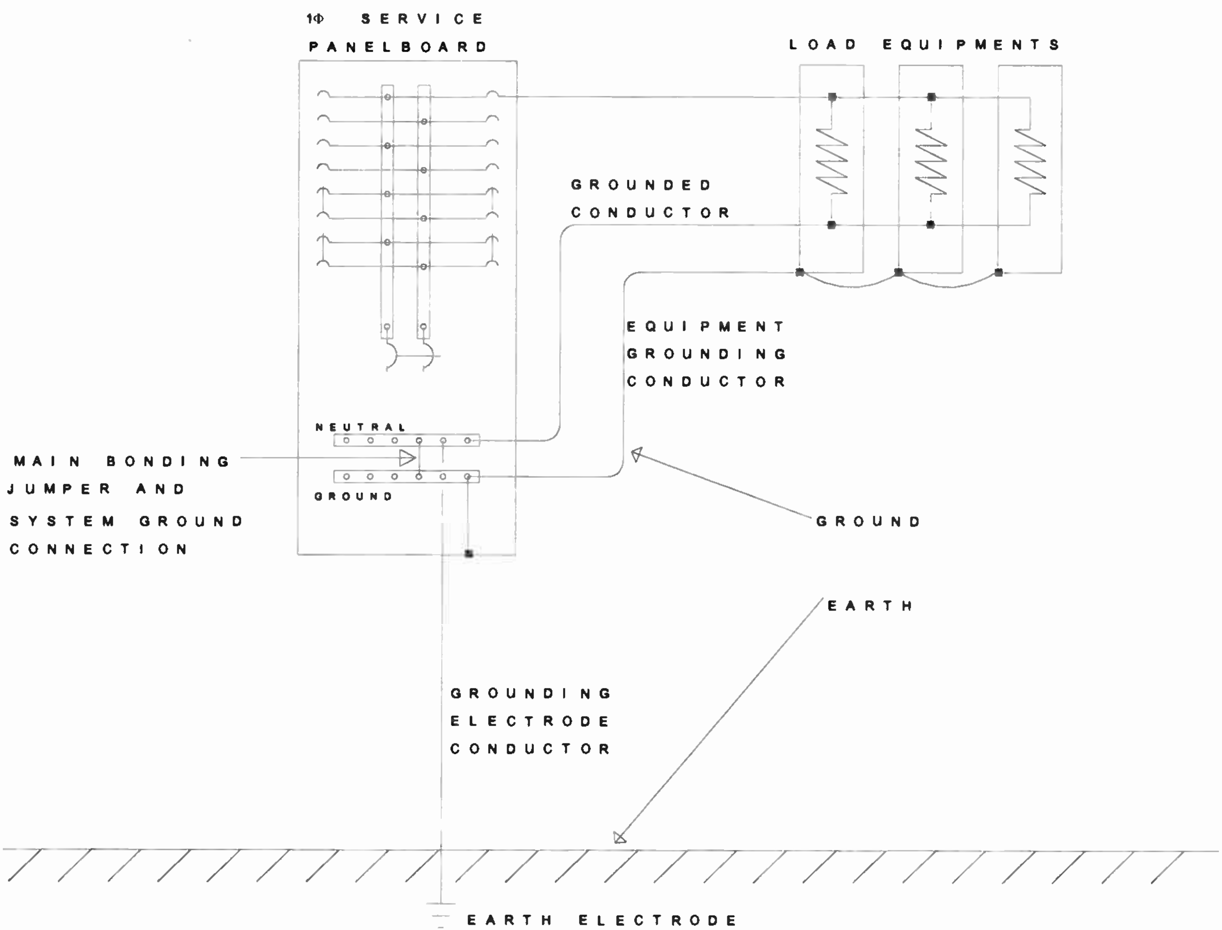


Figure 8.3-1. Equipment grounding and system grounding illustrated.

of the current-carrying conductors of the electrical distribution system to ground. The Figure shows the neutral conductor intentionally connected to ground and the earth. This conductor is called the *grounded* conductor because it is intentionally grounded. The purpose of system grounding is overvoltage control and equipment safety through fault isolation. An ungrounded system is subject to serious overvoltages under conditions such as intermittent ground faults, resonant conditions and contact with higher voltage systems. Fault isolation is achieved by providing a low impedance return path from the load back to the source which will ensure operation of overcurrent devices in the event of a ground fault. The system ground connection makes this possible by connecting the equipment grounding system to the low side of the voltage source. Methods of system grounding include solidly grounded, ungrounded and impedance grounded.

Solidly grounded means that an intentional zero impedance connection is made between a current-carrying conductor and ground. The single-phase (1Φ) system shown in Figure 8.3-1 is solidly grounded. A solidly grounded, three-phase, four-wire, wye system is illustrated in Figure 8.3-2. The neutral is connected directly to ground with no impedance installed in the neutral circuit. The NEC permits this connection to be made at the service only. The advantages of a solidly grounded wye system include reduced magnitude of transient overvoltages, improved fault protection and faster location of ground faults. There is one disadvantage of the solidly grounded wye system. For low level arcing ground faults, the application of sensitive, properly coordinated, ground fault protection (GFP) devices is necessary to prevent equipment damage from arcing ground faults. The NEC requires arcing ground fault protection at 480Y/277V services, and a maximum sensitivity limit of 1200 A is permitted. Severe damage is less frequent at the lower voltage 208 V systems, where the arc may be self extinguishing.

Ungrounded means that there is no intentional connection between a current-carrying conductor and ground. However, charging capacitance will create unintentional capacitive coupling from each phase to ground making the system essentially a capacitance grounded system. A three-phase, three-wire system from an ungrounded delta source is illustrated in Figure 8.3-3. The most important advantage of an ungrounded system is that an accidental ground fault in one phase

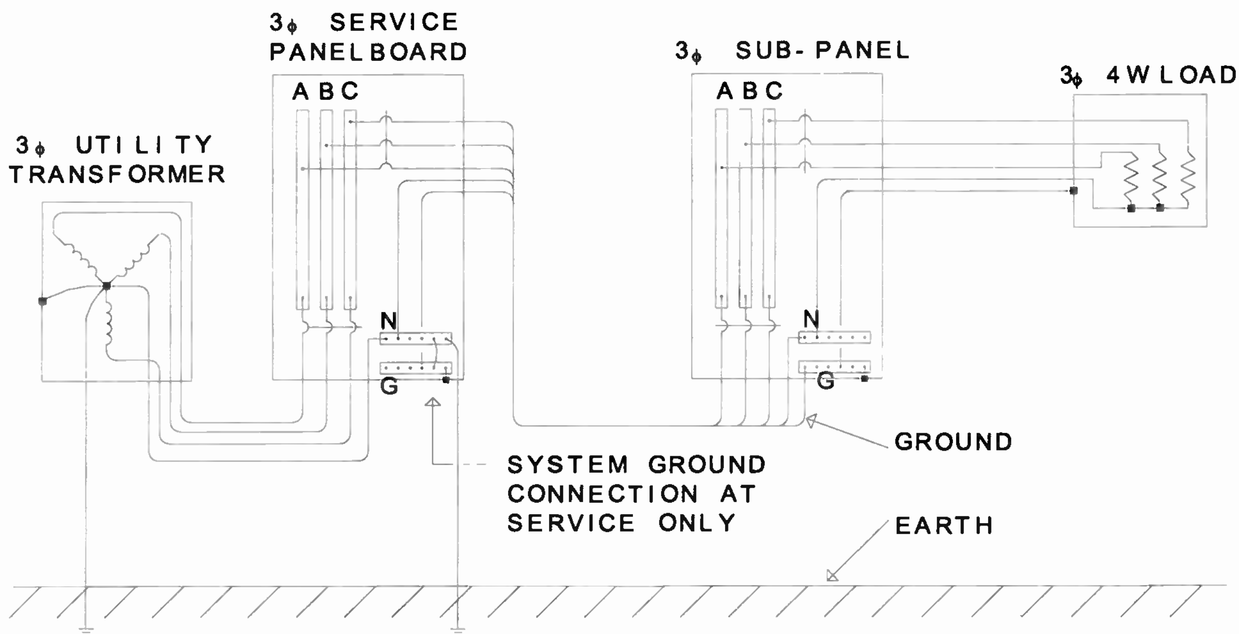


Figure 8.3-2. Solidly grounded wye system.

does not require immediate removal. This allows for continuity of service which made the ungrounded delta system very popular in the past. However, ungrounded systems have serious disadvantages. Since there is no fixed system ground point, it is difficult to locate the first ground fault and to sense the magnitude of fault current.

As a result, the fault is often permitted to remain on the system for a long time. If a second fault should occur before the first one is removed, and the second fault is on a different phase, the result will be a double line-to-ground fault causing serious arcing damage. Another problem with the ungrounded delta system is the occurrence of high transient overvoltages from

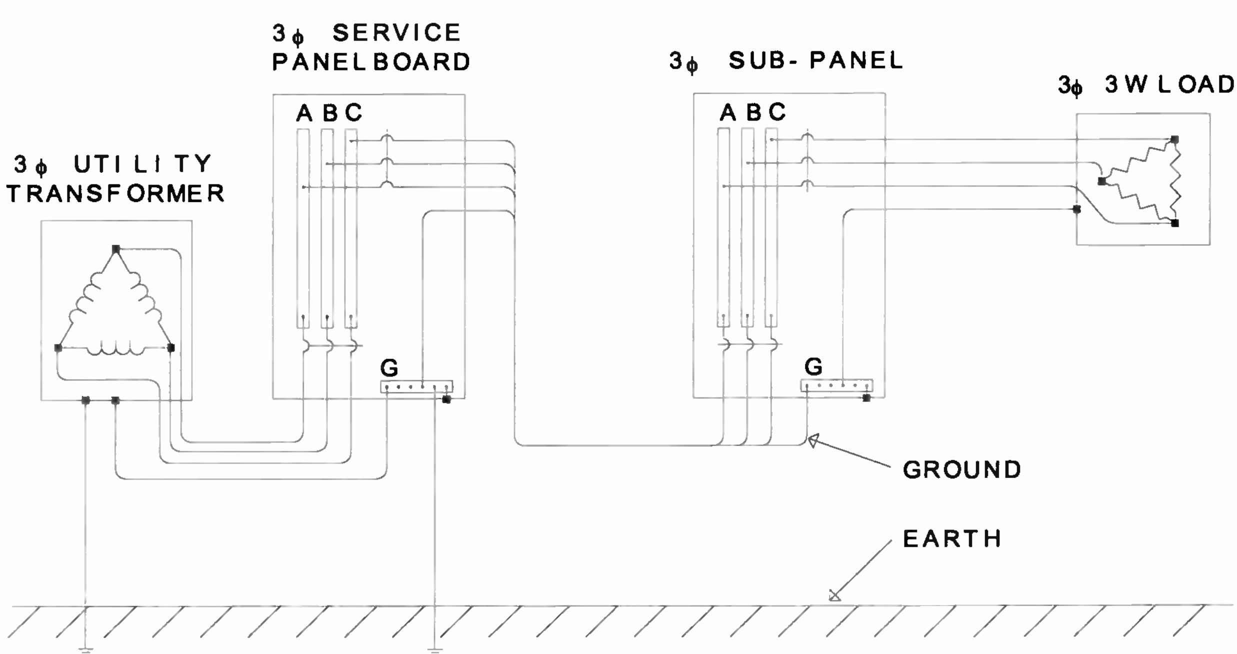


Figure 8.3-3. Ungrounded delta system.

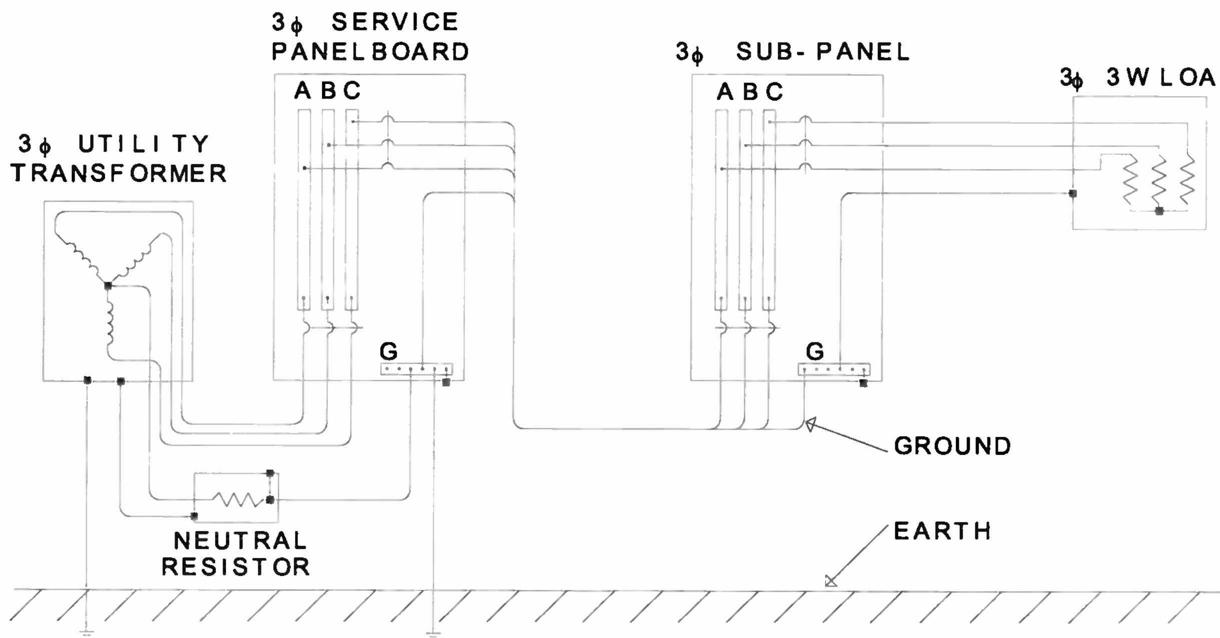


Figure 8.3-4. High resistance grounded wye system.

phase-to-ground. Transient overvoltages are caused by intermittent ground faults, with overvoltages capable of reaching a phase-to-ground voltage of from six to eight times the phase-to-neutral voltage. Sustained overvoltages may ultimately result in insulation failure and thus more ground faults.

Impedance grounded means that an intentional impedance connection is made between a current-carrying conductor and ground. The *high resistance grounded* wye system, illustrated in Figure 8.3-4, is an alternative to solidly grounded and ungrounded systems. High resistance grounding will limit ground fault current to a few amperes, thus removing the potential for arcing damage inherent in solidly grounded systems. The ground reference point is fixed, and relaying methods can locate first faults before damages from second faults occur. Internally generated transient overvoltages are reduced since the neutral to ground resistor dissipates any charge which may buildup on system charging capacitance. Specific requirements for impedance-grounded systems are listed in NEC Article 250.

Table 8.3-1 compares the three most common methods of system grounding. There is no one *best* system grounding method for all applications. In choosing among the various options, the designer must consider the requirements for safety, continuity of service and cost. Generally, low voltage systems should be operated solidly grounded. For applications involving continuous processes in industrial plants or where shutdown might create a hazard, a high resistance grounded wye system, or a solidly grounded wye system with an alternate power supply, may be used. The high resistance grounded wye system combines many of the advantages of the ungrounded delta system and the solidly grounded wye system. IEEE Standard 142 suggests that medium voltage systems less than 15 kV be low resistance grounded to limit ground fault damage, yet permit sufficient current for detection and isolation of ground faults. IEEE Standard 142 also suggests that medium voltage systems over 15 kV be solidly grounded. Solid grounding should include sensitive ground fault relaying in accordance with the NEC.

Table 8.3-1
Comparison of System Grounding Methods.

	System Grounding Method		
Characteristic Assuming No Fault Escalation	Solidly-Grounded	Un-Grounded	High Resistance
Operation of Overcurrent Device on First Ground Fault	Yes	No	No
Control of Internally-Generated Transient Overvoltages	Yes	No	Yes
Control of Steady-State Overvoltages	Yes	No	Yes
Flash Hazard	Yes	No	No
Equipment Damage from Arcing Ground-Faults	Yes	No	No
Overvoltage (on Unfaulted Phases) from Ground-Fault	L-N Voltage	>> L-L Voltage	L-L Voltage
Can serve line-to-neutral loads	Yes	No	No

Earth Connections

The process of connecting the grounding system to the earth is called *earthing*, and consists of immersing a metal electrode or system of electrodes into the earth. The conductor which connects the grounding system to earth is called the *grounding electrode conductor*. The grounding electrode conductor is illustrated in Figure 8.3-1. Its function is to keep the entire grounding system at earth potential (voltage equalization during lightning and other transients) rather than for conducting ground fault current. Therefore, the NEC allows reduced sizing requirements for that portion of the grounding electrode conductor when connected to *made* electrodes (see NEC Article 250 for exact requirements).

The basic measure of effectiveness of an earth electrode system is called *earth electrode resistance*. Earth electrode resistance is the resistance, in ohms, between the point of connection and a distant point on the earth called *remote earth*. Remote earth, about 25 ft from the driven electrode, is the point where earth electrode resistance does not increase appreciably when this distance is increased. Earth electrode resistance consists of the sum of the resistance of the metal electrode (negligible) plus the contact resistance between the electrode and the soil (negligible) plus the soil resistance itself. Thus, for all practical purposes, earth electrode resistance equals the soil resistance. The soil resistance is nonlinear, with most of the earth resistance contained within several feet of the electrode. Furthermore, current flows only through the electrolyte portion of the soil, not the soil itself. Thus, soil resistance varies as the electrolyte content (moisture and salts) of the soil varies. Without electrolyte, soil resistance would be very large.

Soil resistance is a function of soil resistivity. A one cubic meter sample of soil with a resistivity (ρ) of 1 Ω-meter will present a resistance (R) of 1 Ω between opposite faces. A broad variation of soil resistivity occurs as a function of soil types, and soil resistivity can be estimated or measured directly. Soil resistivity is usually measured by injecting a known current into a given volume of soil and measuring the resulting voltage drop. When soil resistivity is known, the earth electrode resistance of any given configuration (single rod, multiple rods, or ground ring) may be determined by using standard equations developed by Sunde, Schwarz, and others[5,6]. Three standard grounding equations are listed.

$$R_1 = \frac{0.52\rho}{lr}\left(\ln\frac{96lr}{Dr} - 1\right) \qquad (1)$$

Where:

R_1 = Resistance in ohms of a single driven ground rod

ρ = Soil resistivity in ohm-meters

lr = Length of ground rod in feet

Dr = Diameter of ground rod in inches

$$R_N = \frac{1}{N}\left[R_1 + \frac{1.05\rho}{S}\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}\right)\right] \qquad (2)$$

Where:

R_N = Resistance in ohms of N driven ground rods

N = Number of equally spaced ground rods

R_1 = Resistance in ohms of a single driven ground rod

S = Uniform Spacing of ground rods in feet

$$R_{GR} = \frac{0.52\rho}{(N)(lr)}\left[\ln\left(\frac{96lr}{Dr}\right) - 1 + \frac{2K_1 lr}{\sqrt{A}}(\sqrt{N} - 1)^2\right] \qquad (3)$$

Where:

R_{GR} = Resistance in ohms of ground ring configuration

A = Area of coverage at farthest dimensions in square feet

K_1 = Length-to-width ratio coefficient

Length-to-Width Ratio	$\approx K_1$
1	1.37
2	1.31
3	1.26
4	1.22
5	1.18
6	1.15
7	1.12
8	1.10

Earth resistance values should be as low as practicable, but are a function of the application. The NEC approves the use of a single made electrode if the earth resistance does not exceed 25 Ω. IEEE Standard 1100 reports that the very low earth resistance values specified for computer systems in the past is not necessary. Methods of reducing earth resistance values include the use of multiple electrodes in parallel, ground rings, increased ground rod lengths, installation of ground rods to the permanent water level, increased area of coverage of ground rings and use of concrete-encased electrodes, ground wells, and electrolytic electrodes.

EARTH ELECTRODES

Earth electrodes may be *made* electrodes, *natural* electrodes, or *special purpose* electrodes. Made electrodes include driven rods, buried conductors, ground mats, buried plates and ground rings. The electrode selected is a function of the type of soil and the available depth. Driven electrodes are used where bedrock is 10 ft or more below the surface. Mats or buried conductors are used for lesser depths. Buried plates are not widely used because of the higher cost when compared to rods.

[5,6] Sunde, E.D., *Earth Conduction Effects in Transmission Systems*, Von Nonstrand Co., 1949. Schwarz, S.J., *Analytical Expression for Resistance of Grounding Systems*, AIEE Transactions Vol. 73, Part III-B, pp 1011–1016, 1954.

Ground rings employ equally spaced driven electrodes interconnected with buried conductors. Ground rings are used around large buildings, around small unit substations, and in areas having high soil resistivity.

Natural electrodes include buried water pipe electrodes and concrete encased electrodes. The NEC lists underground metal water piping, available on the premises and not less than 10 ft in length, as part of a preferred grounding electrode system. Since the use of plastic pipes in new water systems will impair the effectiveness of water pipe electrodes, the NEC requires that metal underground water piping be supplemented by an additional approved electrode. Concrete below ground level is a good electrical conductor. Thus, metal electrodes encased in such concrete will function as excellent grounding electrodes. The application of concrete encased electrodes is covered in IEEE Standard 142.

A special purpose electrode called the *electrolytic* or *chemically charged electrode*, consists of a ground rod installed in a canister containing an electrolytic salt. The salt absorbs moisture from the atmosphere through *breather* holes at the top of the cannister. As the salt dissolves in the moisture, it forms a homogeneous electrolytic solution. Gravity and changes in atmospheric pressure cause the electrolytic solution to leak out through *weep* holes at the bottom of the cannister. As the electrolytic solution soaks into the surrounding soil, it reduces the soil resistivity and also helps to hold it relatively constant. The result is a consistently lowered and uniform earth electrode resistance.

FACILITIES GROUNDING FOR COMPUTER AND ELECTRONICS EQUIPMENT

Noise control is an important aspect of computer and electronic systems. The process of noise control through proper grounding techniques is called *referencing*. For this discussion, electronic systems will be viewed as a multiplicity of signal sources transmitting signals to a multiplicity of loads. The ideal electronic circuit consists of a simple signal source supplying a load via a pair of leads as shown in Figure 8.3-5.

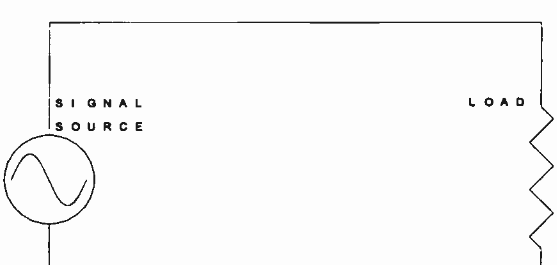


Figure 8.3-5. Ideal electronic circuit.

This source load electronic pair is ideal and free of interference because the impedance of the signal return path is zero and the signal return is a dedicated path.

Practically speaking, however, the impedance of the signal return path is not zero and dedicated return paths for each source load pair are not practical. Packaged electronics systems typically incorporate a common signal reference plane which serves as a common return path for numerous source load pairs (see to Figure 8.3-6(a)). The signal reference plane may be a large dedicated area on a printed circuit board, the metal chassis or enclosure of the electronic equipment, or the metal frame or mounting rack that houses several different units. Ideally, the signal reference plane offers zero impedance to the signal current. Practically, however, the signal reference plane has a finite impedance. The practical result is called *common impedance* or *conductive coupling* (see Figure 8.3-6(b)). Since a practical signal reference plane has a finite impedance, current flow in the plane will produce potential differences between various points on the plane. Source load pairs referenced to the plane will, therefore, experience interference as a result. Z_R is common to both circuits referenced to the plane in Figure 8.3-6(b). Thus, I_1 and I_2 returning to their respective sources will produce interference voltages by flowing through Z_R. The total interference voltage drop felt across Z_R causes the source reference A to be at a different potential than the load reference B. This difference in potential is often called *ground voltage shift* (even though ground may not even be involved), and is a major source of noise and interference in electronic circuits.

Ground voltage shifts can also be caused by electromagnetic or electrostatic fields in close proximity to the source load pairs. The interference source induces interference voltages into any closed loop by antenna action. This loop is called a *ground loop* (even though ground may not be involved). Interference voltages can be minimized by reducing the loop area as much as possible. This may be very difficult if the loop includes an entire room. The interference voltage can be eliminated entirely by breaking the loop.

Within individual electronic equipments, the signal reference plane consists of a metal plate or the metal enclosure or a rack assembly as previously discussed. Between units of equipment that are located in different rooms on different floors or even in different buildings, the signal reference planes of each unit must be connected together via interconnected wiring such as coax shields or separate conductors. This action, of course, increases the impedance between signal reference planes and makes noise control more difficult. Reducing noise caused by common impedance coupling is a matter of reducing the impedance of the interconnected signal reference planes. Regardless of the configuration encountered (circuit board, individual electronic equipment or equipments remotely located within a facility or in separate buildings), the next question to be answered is should the signal reference be connected to ground? Floating signal grounding, single point grounding and multipoint grounding are methods of

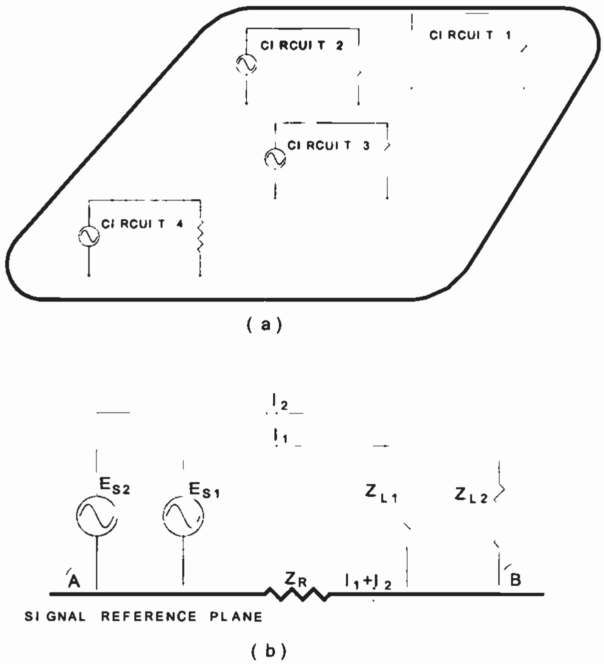


Figure 8.3-6. Equipotential plane.

accomplishing this signal reference to ground connection.

IEEE Standard 1100 recommends multipoint grounding and a signal reference plane for most computer and electronics installations. An ideal equipotential signal reference plane is one which has 0 V (thus zero impedance) between any two points on the plane. Since an *ideal* equipotential plane is not attainable, a *nominal* equipotential plane is accepted. Multipoint grounding connections are made to the plane which ensures minimum ground voltage shift between signal reference systems connected to the plane. Collectively, the current flow in an equipotential plane may be quite large. Between any two equipments, however, the current flow should be low due to the many current paths available.

Practical versions of an equipotential plane include bolted stringer system of a raised computer floor, flat copper strips bonded together at 2 ft centers, copper conductors bonded together at 2 ft centers and single or multiple, flat copper strips connected between equipments.

ISOLATED GROUNDING

Isolated grounding schemes, where the signal reference plane is isolated from equipment ground but connected to an *isolated* electrode in the earth, do not work, are unsafe and violate the NEC. It is thought by some people that the isolated earth connection is *clean* since there is no connection between it and the *dirty* system ground connection at the service entrance. The clean, isolated earth connection is also viewed (incorrectly) as a point where noise currents can flow into the earth and be dissipated. *Kirchoff's Current Law* teaches that any current flowing into the isolated ground must return to the source through another earth connection. Current cannot be dissipated. It must always return to its source. Even lightning current is not dissipated into the earth. It must have a return path (the electrostatic and electromagnetic fields which created the charge buildup and the lightning strike in the first place).

Consider what might happen if the previously described system above is subjected to a lightning strike. Assume that a transient current of 2,000 A flows into the earth and through an earth resistance of 5 Ω between the system ground electrode and the isolated electrode. A more realistic resistance would be even higher, maybe 25 Ω. Two thousand amperes flowing through 5 Ω results in a voltage drop or transient potential of 10,000 V between the two electrodes. Since this potential is impressed between the equipment frame (system ground electrode) and the signal reference plane (isolated electrode), it could result in equipment damage and personnel hazard. Dangerous potential differences between grounding subsystems can be reduced by bonding together all earth electrodes on a facility. A fine print note (FPN) in NEC Article 250 states that the bonding together of separate grounding electrodes will limit potential differences between them and their associated wiring systems.[7]

A facility ground system, then, can be defined as an electrically interconnected system of multiple conducting paths to the earth electrode or system of electrodes. The facility grounding system includes all electrically interconnected grounding subsystems such as the equipment grounding subsystem, the signal reference subsystem, the fault protection subsystem and the lightning protection subsystem.

Isolated ground (IG) receptacles, which are a version of single-point grounding, are permitted by the NEC. Proper application of IG receptacles is very important. They must be used with an insulated equipment grounding conductor, not a bare conductor. Also, only metallic conduit should be used.

SEPARATELY-DERIVED SYSTEMS

A *separately derived system* is a premises wiring system whose power is derived from generator, transformer or converter windings and that has no direct electrical connection, including a solidly connected grounded circuit conductor, to supply conductors originating in another system.[5] Solidly grounded, wye connected, isolation transformers used to supply power to computer room equipment are examples of separately derived systems. Specific requirements for the bonding and grounding of seperately derived systems are listed in NEC Article 250. Other examples of separately derived systems include generators and UPS systems.

[7] NFPA Std 70, *The National Electrical Code*, The National Fire Protection Association, Inc., Quincy, MA, 1996.

KEY TERMS

Grounded Conductor. A system or circuit conductor that is intentionally grounded.

Equipment Grounding Conductor. The conductor used to connect the noncurrent-carrying metal parts of equipment, raceways, and other enclosures to the system grounded conductor, the grounding electrode conductor, or both, at the service equipment or at the source of a separately derived system.

Main Bonding Jumper. The connection between the grounded circuit conductor and the equipment grounding conductor at the service.

Grounding Electrode Conductor. The conductor used to connect the grounding electrode to the equipment grounding conductor, to the grounded conductor, or to both, of the circuit at the service equipment or at the source of a separately derived system.

Service. The conductors and equipment for delivering energy from the electricity supply system to the wiring system of the premises served.

Service Conductors. The supply conductors that extend from the street main or from transformers to the service equipment of the premises supplied.

Service Equipment. The necessary equipment, usually consisting of a circuit breaker or switch and fuses, and their accessories, located near the point of entrance of supply conductors to a building or other structure, or an otherwise defined area, and intended to constitute the main control and means of cutoff of the supply.

Equipotential Plane. A mass of conducting material which offers a negligible impedance to current flow, thus producing zero volts (equipotential) between points on the plane.

Floating Signal Grounding. A non-grounding system in which all electronic signal references are isolated from ground.

Single-Point Grounding. A grounding system in which all electronic signal references are bonded together and grounded at a single point.

Multipoint Grounding. A grounding system in which all electronic signal references are grounded at multiple points.

FURTHER INFORMATION

Ground Fault Protection for Solidly Grounded Low Voltage Systems, General Electric Bulletin EESG II-AP-7, December, 1974.

NEMA PB2.2, *Application Guide for Ground Fault Protective Devices for Equipment*, National Electrical Manufacturers Association, Washington DC, current edition.

Lazar, Irwin, *System Grounding in Industrial Power Systems*, Specifying Engineer, May 1978–Jan 1979.

Techniques of Electrical Design and Construction (Vol 4)—Protecting Against Ground Faults, McGraw-Hill, 1979.

Kaufmann, R.H., *Some Fundamentals of Equipment Grounding Circuit Design,* GEB #957, Industrial Power Systems Applications, General Electric Corporation.

8.4
STANDBY POWER SYSTEMS

JERRY WHITAKER
EDITOR-IN-CHIEF

INTRODUCTION

When utility company power problems are discussed, most people immediately think of blackouts. The lights go out, and everything stops. With the facility down and in the dark, there is nothing to do but sit and wait until the utility company finds the problem and corrects it. This process generally takes only a few minutes. There are times, however, when it can take hours. In some remote locations, it can even take days.

Blackouts are, without a doubt, the most troublesome utility company problem that a broadcast facility will have to face. Statistics show that power failures are, generally speaking, a rare occurrence in most areas of the country and are short in duration. Studies have shown that 50% of blackouts last six seconds or less, and 35% are less than 11 minutes long. These failure rates usually are not cause for concern to commercial users, except where computer based operations, transportation control systems, medical facilities, and communications sites—including radio and television stations—are concerned.

When continuity of operation is critical, redundancy must be carried throughout the system. The site never should depend upon one critical path for ac power. For example, if the facility is fed by a single step-down transformer, a lightning flash or other catastrophic event could result in a transformer failure that would bring down the entire site. A replacement could take days or even weeks.

Blackout Effects

A broadcast facility that is down for even five minutes can suffer a significant loss of audience that may take hours or days to rebuild. A blackout affecting a transportation or medical center could be life threatening. Coupled with this threat is the possibility of extended power service loss due to severe storm conditions. Many broadcast and communications relay sites are located in remote, rural areas or on mountaintops. Neither of these kinds of locations are well known for their power reliability. It is not uncommon in mountainous areas for utility company service to be out for extended periods after a major storm. Few operators are willing to take such risks with their business. Most choose to install standby power systems at appropriate points in the equipment chain.

The cost of standby power for a facility can be substantial, and an examination of the possible alternatives should be conducted before any decision on equipment is made. Management must clearly define the direct and indirect costs and weigh them appropriately. Include the following items in the cost versus risk analysis:

- Standby power system equipment purchase and installation cost
- Exposure of the system to utility company power failure
- Alternative operating methods available to the facility
- Direct and indirect costs of lost uptime because of blackout conditions.

A distinction must be made between *emergency* and *standby* power sources. Strictly speaking, emergency systems supply circuits legally designated as being essential for safety to life and property. Standby power systems are used to protect a facility against the loss of productivity resulting from a utility company power outage.

STANDBY POWER OPTIONS

To ensure the continuity of ac power, many commercial/industrial facilities depend upon either two separate utility services or one utility service plus on-site generation. Because of the growing complexity of electrical systems, attention must be given to power supply reliability.

The engine generator shown in Figure 8.4-1 is the classic standby power system. An automatic transfer switch monitors the ac voltage coming from the utility company line for power failure conditions. Upon detection of an outage for a predetermined period of time (generally 1 to 10 seconds), the standby generator is started; after the generator is up to speed, the load is transferred from the utility to the local generator. Upon return of the utility feed, the load is switched back, and the generator is stopped. This basic type of system is used widely in industry and provides economical protection against prolonged power outages (5 minutes or more).

The transfer device shown in Figure 8.4-1 is a contactor type, break-before-make unit. By replacing the simple transfer device shown with an automatic over-

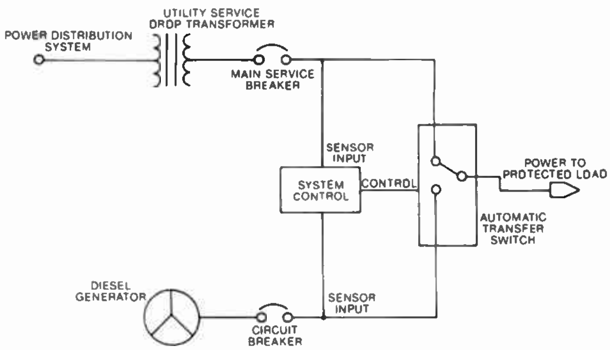


Figure 8.4-1. The classic standby power system using an engine-generator set. This system protects a facility from prolonged utility company power failures.

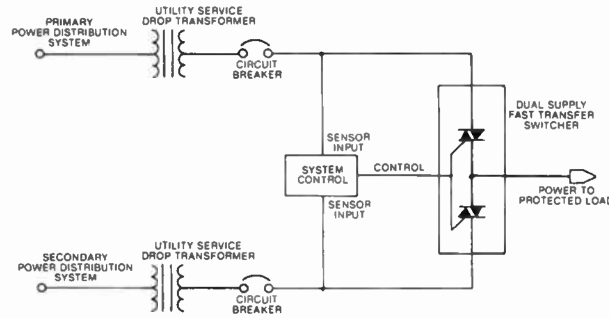


Figure 8.4-3. The dual utility feeder system of ac power loss protection. An automatic transfer switch changes the load from the main utility line to the standby line in the event of a power interruption.

lap (*static*) transfer switch, as shown in Figure 8.4-2, additional functionality can be gained. As described in Chapter 8.2, *AC Power Conditioning*, the overlap transfer switch permits the on-site generator to be synchronized with the load, making a clean switch from one energy source to another. This functionality offers the following benefits:

- Switching back to the utility feed from the generator can be accomplished without interruption in service
- The load can be cleanly switched from the utility to the generator in anticipation of utility line problems (such as an approaching severe storm)
- The load can be switched to and from the generator to accomplish *load* shedding objectives.

Dual Feeder System

In some areas, usually metropolitan centers, two utility company power drops can be brought into a facility as a means of providing a source of standby power. As shown in Figure 8.4-3, two separate utility service drops—from separate power distribution systems—are brought into the plant, and an automatic transfer switch changes the load to the backup line in the event of a main line failure. The dual feeder system provides an advantage over the auxiliary diesel arrangement in that power transfer from main to standby can be made in a fraction of a second if a static transfer switch is used. Time delays in the diesel generator system limit its usefulness to power failures lasting more than several minutes.

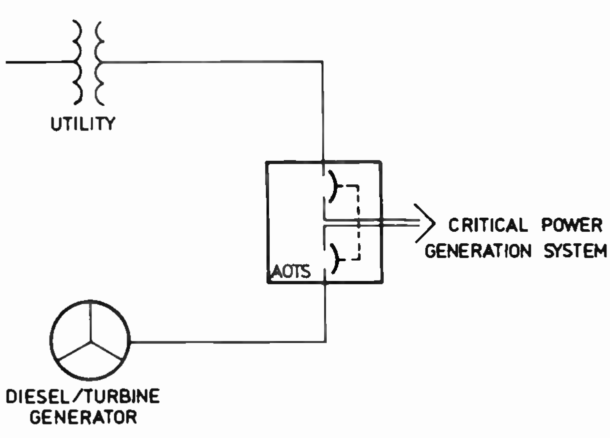


Figure 8.4-2. The use of a static transfer switch to transfer the load from the utility company to the on-site generator.

The dual feeder system of protection is based on the assumption that each of the service drops brought into the facility is routed via different paths. This being the case, the likelihood of a simultaneous failure on both power lines is remote. The dual feeder system will not, however, protect against area-wide power failures, which may occur from time to time.

The dual feeder system is limited primarily to urban areas. Rural or mountainous regions generally are not equipped for dual redundant utility company operation. Even in urban areas, the cost of bringing a second power line into a facility can be high, particularly if special lines must be installed for the feed. If two separate utility services are available at or near the site, redundant feeds generally will be less expensive than engine driven generators of equivalent capacity.

Figure 8.4-4 illustrates a dual feeder system that utilizes both utility inputs simultaneously at the facility. Notice that during normal operation, both ac lines feed loads, and the *tie* circuit-breaker is open. In the event of a loss of either line, the circuit-breaker switches reconfigure the load to place the entire facility on the single remaining ac feed. Switching is performed automatically; manual control is provided in the event of a planned shutdown on one of the lines.

Peak Power Shaving

Figure 8.4-5 illustrates the use of a backup diesel generator for both standby power and *peak power shaving* applications. Commercial power customers often can realize substantial savings on utility company bills by reducing their energy demand during certain hours of the day. An automatic overlap transfer switch is used to change the load from the utility company system to the local diesel generator. The changeover is accomplished by a static transfer switch that does not disturb the operation of load equipment. This application of a standby generator can provide financial return to the facility, whether or not the unit is ever needed to carry the load through a commercial power failure.

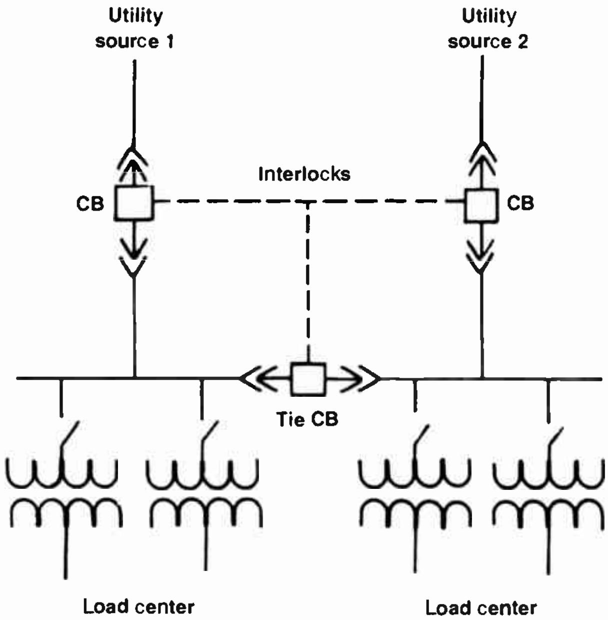


Figure 8.4-4. A dual utility feeder system with interlocked circuit breakers.

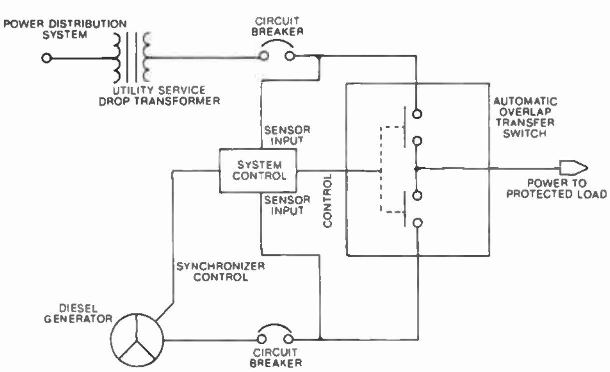


Figure 8.4-5. The use of a diesel generator for standby power and peak power shaving applications. The automatic overlap (static) transfer switch changes the load from the utility feed to the generator instantly so that no disruption of normal operation is encountered.

Advanced System Protection

A more sophisticated power control system is shown in Figure 8.4-6, where a dual feeder supply is coupled with a motor generator (m-g) set to provide clean, undisturbed ac power to the load. The m-g set will smooth over the transition from the main utility feed to the standby, often making a commercial power failure unnoticed by on-site personnel. As discussed in Chapter 8.2, a conventional m-g typically will give up to 1/2 seconds of power fail ride-through, more than enough to accomplish a transfer from one utility feed to the other. This standby power system is further refined in the application illustrated in Figure 8.4-7, where a diesel generator has been added to the system. With the automatic overlap transfer switch shown at

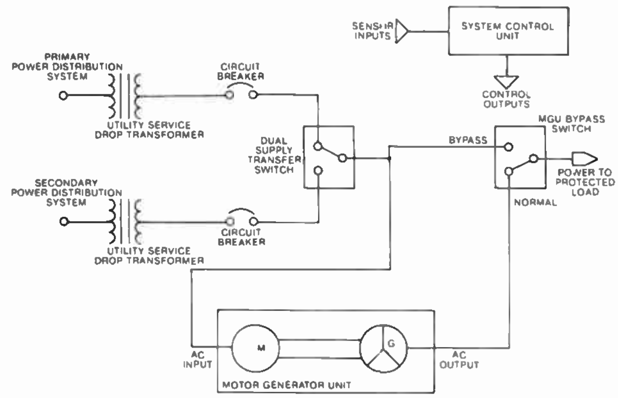


Figure 8.4-6. A dual feeder standby power system using a motor-generator set to provide power fail ride-through and transient-disturbance protection. Switching circuits allow the m-g set to be bypassed, if necessary.

the generator output, this arrangement also can be used for peak demand power shaving.

Figure 8.4-8 shows a simplified schematic diagram of a 220 kW uninterruptible power system (UPS) utilizing dual utility company feed lines, a 750 kVA gas engine-generator, and five dc-driven m-g sets with a 20 minute battery supply at full load. The five m-g sets operate in parallel, each is rated for 100 kW output. Only three are needed to power the load, but four are on-line at any given time. The fifth machine provides redundancy in the event of a failure or for scheduled maintenance work. The batteries are always on-line under a slight charge across the 270 V dc bus. Two separate natural gas lines, buried along different land routes, supply the gas engine. Local gas storage capacity also is provided.

Choosing a Generator

Engine-generator sets are available for power levels ranging from less than 1 kVA to several thousand kVA

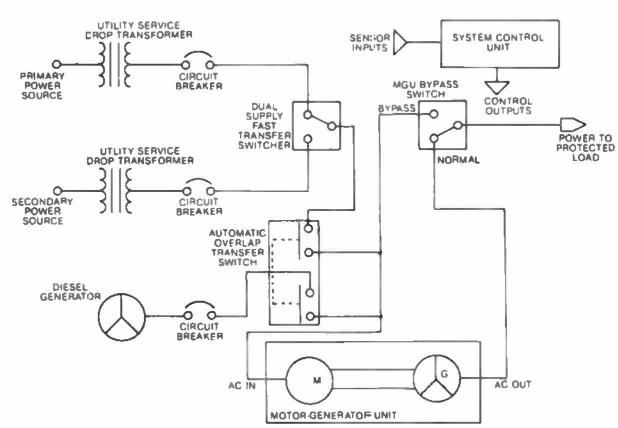


Figure 8.4-7. A premium power supply backup and conditioning system using dual utility company feeds, a diesel generator, and a motor-generator set. An arrangement such as this would be used for critical loads that require a steady supply of clean ac.

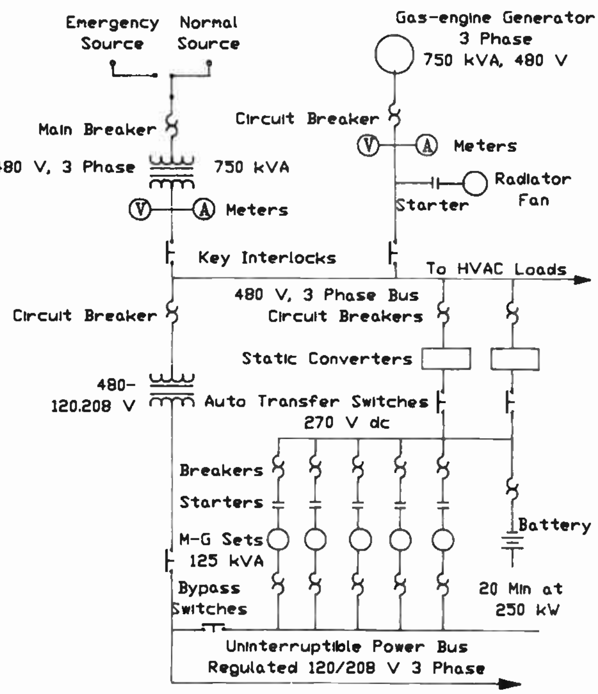


Figure 8.4-8. Simplified installation diagram of a high reliability power system incorporating dual utility feeds, a standby gas engine-generator and five battery-backed dc m-g sets.[1]

or more. Machines also may be paralleled to provide greater capacity. Engine-generator sets typically are classified by the type of power plant used:

- **Diesel.** Advantages: rugged and dependable, low fuel costs, low fire and/or explosion hazard. Disadvantages: somewhat more costly than other engines, heavier in smaller sizes
- **Natural and liquefied petroleum gas.** Advantages: quick starting after long shutdown periods, long life, low maintenance. Disadvantage: availability of natural gas during area-wide power failure subject to question
- **Gasoline.** Advantages: rapid starting, low initial cost. Disadvantages: greater hazard associated with storing and handling gasoline, generally shorter mean time between overhaul
- **Gas turbine.** Advantages: smaller and lighter than piston engines of comparable horsepower, rooftop installations practical, rapid response to load changes Disadvantages: longer time required to start and reach operating speed, sensitive to high input air temperature.

The type of power plant chosen is usually determined primarily by the environment in which the system will be operated and by the cost of ownership. For example, a standby generator located in an urban area office complex may be best suited to the use of an engine powered by natural gas, because of the problems inherent in storing large amounts of fuel. State or local building codes may place expensive restrictions on fuel storage tanks and make the use of a gasoline or diesel powered engine impractical. The use of propane usually is restricted to rural areas. The availability of propane during periods of bad weather (when most power failures occur) also must be considered.

The generator rating for a standby power system should be chosen carefully and should take into consideration the anticipated future growth of the plant. It is good practice to install a standby power system rated for at least 25% greater output than the current peak facility load. This headroom gives a margin of safety for the standby equipment and allows for future expansion of the facility without overloading the system.

An engine driven standby generator typically incorporates automatic starting controls, a battery charger and automatic transfer switch (see Figure 8.4-9). Control circuits monitor the utility supply and start the engine when there is a failure or a sustained voltage drop on the ac supply. The switch transfers the load as soon as the generator reaches operating voltage and frequency. Upon restoration of the utility supply, the switch returns the load and initiates engine shutdown. The automatic transfer switch must meet demanding requirements, including:

- Carrying the full rated current continuously
- Withstanding fault currents without contact separation
- Handling high inrush currents
- Withstanding many interruptions at full load without damage.

The nature of most power outages requires a sophisticated monitoring system for the engine-generator set. Most power failures occur during periods of bad weather. Most standby generators are unattended. More often than not, the standby system will start, run and shut down without any human intervention or supervision. For reliable operation, the monitoring system must check the status of the machine continually to ensure that all parameters are within normal limits.

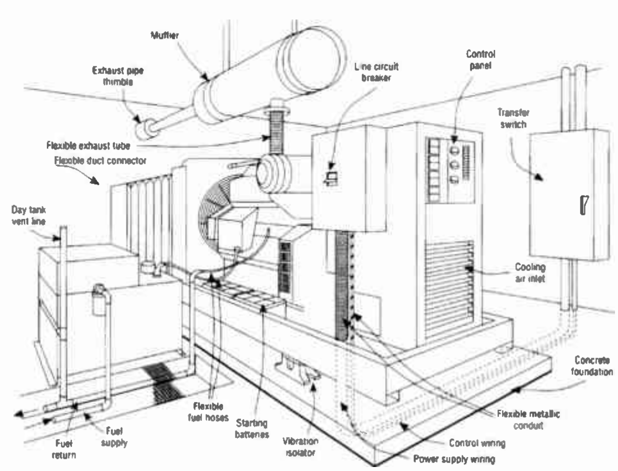


Figure 8.4-9. Typical configuration of an engine-generator set.[2] (*Used with permission.*)

Time-delay periods usually are provided by the controller that require an outage to last from 5 to 10 seconds before the generator is started and the load is transferred. This prevents false starts that needlessly exercise the system. A time delay of 5 to 30 minutes usually is allowed between the restoration of utility power and return of the load. This delay permits the utility ac lines to stabilize before the load is reapplied.

The transfer of motor loads may require special consideration, depending upon the size and type of motors used at a plant. If the residual voltage of the motor is out of phase with the power source to which the motor is being transferred, serious damage may result to the motor. Excessive current draw also may trip overcurrent protective devices. Motors above 50 hp with relatively high load inertia in relation to torque requirements, such as flywheels and fans, may require special controls. Restart time delays are a common solution.

Automatic starting and synchronizing controls are used for multiple engine-generator installations. The output of two or three smaller units can be combined to feed the load. This capability offers additional protection for the facility in the event of a failure in any one machine. As the load at the facility increases, additional engine-generator systems can be installed on the standby power bus.

Generator Types

Generators for standby power applications may be *induction* or *synchronous* machines. Most engine-generator systems in use today are synchronous because of the versatility, reliability and capability of operating independently that this approach provides. Most modern synchronous generators are of the *revolving field alternator* design. Essentially, this means that the armature windings are held stationary and the field is rotated. Therefore, generated power can be taken directly from the stationary armature windings. Revolving armature alternators are less popular because the generated output power must be derived via slip rings and brushes.

The exact value of the ac voltage produced by a synchronous machine is controlled by varying the current in the dc field windings, while frequency is controlled by the speed of rotation. Power output is controlled by the torque applied to the generator shaft by the driving engine. In this manner, the synchronous generator offers precise control over the power it can produce.

Practically all modern synchronous generators use a *brushless exciter.* The exciter is a small ac generator on the main shaft; the ac voltage produced is rectified by a three-phase rotating rectifier assembly also on the shaft. The dc voltage thus obtained is applied to the main generator field, which is also on the main shaft. A voltage regulator is provided to control the exciter field current, and in this manner, the field voltage can be precisely controlled, resulting in a stable output voltage.

The frequency of the ac current produced is dependent on two factors: the number of poles built into the machine and the speed of rotation (rpm). Because the output frequency must normally be maintained within strict limits (60 Hz or 50 Hz), control of the generator speed is essential. This is accomplished by providing precise rpm control of the *prime mover,* which is performed by a governor.

There are many types of governors; however, for auxiliary power applications, the *isochronous governor* is normally selected. The isochronous governor controls the speed of the engine so that it remains constant from no load to full load, assuring a constant ac power output frequency from the generator. A modern system consists of two primary components: an electronic speed control and an actuator that adjusts the speed of the engine. The electronic speed control senses the speed of the machine and provides a feedback signal to the mechanical/hydraulic actuator, which in turn positions the engine throttle or fuel control to maintain accurate engine rpm.

The National Electrical Code provides guidance for safe and proper installation of on-site engine-generator systems. Local codes may vary and must be reviewed during early design stages

UPS Systems

An uninterruptible power system (UPS) is an elegant solution to power outage concerns. The output of the UPS inverter may be a sine wave or pseudo sine wave. When shopping for a UPS system, consider the following:

- Power reserve capacity for future growth of the facility
- Inverter current surge capability (if the system will be driving inductive loads, such as motors)
- Output voltage and frequency stability over time and with varying loads
- Required battery supply voltage and current. Battery costs vary greatly, depending upon the type of units needed
- Type of UPS system—forward-transfer type (where the inverter is normally inactive) or reverse-transfer type (where the inverter is normally active)—required by the particular application. Some sensitive loads may not tolerate even brief interruptions of the ac power source
- Inverter efficiency at typical load levels. Some inverters have good efficiency ratings when loaded at 90% of capacity, but poor efficiency when lightly loaded
- Size and environmental requirements of the UPS system. High power UPS equipment requires a large amount of space for the inverter/control equipment and batteries Battery banks often require special ventilation and ambient temperature control.

Standby Power System Noise

Noise produced by backup power systems can be a serious problem if not addressed properly. Standby generators, m-g sets and UPS systems produce noise that can disturb building occupants and irritate neighbors and/or landlords.

The noise associated with electrical generation usually is related to the drive mechanism, most commonly an internal combustion engine. The amplitude of the noise produced is directly related to the size of the engine-generator set. First consider whether noise reduction is a necessity. Many building owners have elected to tolerate the noise produced by a standby power generator because its use is limited to emergency situations. During a crisis, when the normal source of power is unavailable, most people will tolerate noise associated with a standby generator.

If the decision is made that building occupants can live with the noise of the generator, care must be taken in scheduling the required testing and exercising of the unit. Whether testing occurs monthly or weekly, it should be done on a regular schedule.

If it has been determined that the noise should be controlled, or at least minimized, the easiest way to achieve this objective is to physically separate the machine from occupied areas—this may be easier said than done. Because engine noise is predominantly low-frequency in character, walls and floor/ceiling construction used to contain the noise must be massive. Lightweight construction, even though it may involve several layers of resiliently mounted drywall, is ineffective in reducing low-frequency noise. Exhaust noise is a major component of engine noise but, fortunately, it is easier to control. When selecting an engine-generator set, select the highest quality exhaust muffler available. Such units often are identified as hospital grade mufflers.

Engine-generator sets also produce significant vibration. The machine should be mounted securely to a slab-on-grade or an isolated basement floor, or it should be installed on vibration isolation mounts. Such mounts usually are specified by the manufacturer.

Because a UPS system or m-g set is a source of continuous power, it must run continuously. Noise must be adequately controlled. Physical separation is the easiest and most effective method of shielding occupied areas from noise. Enclosure of UPS equipment usually is required, but noise control is significantly easier than for an engine-generator because of the lower noise levels involved. Nevertheless, the low-frequency 120 Hz fundamental of a UPS system is difficult to contain adequately and massive constructions may be necessary. Vibration control also is required for most UPS and m-g gear.

Batteries

Batteries are the lifeblood of most UPS systems. Important characteristics include the following:

- Charge capacity—how long the battery will operate the UPS
- Weight
- Charging characteristics
- Durability/ruggedness

Additional features that add to the utility of the battery include:

- Built-in status/temperature/charge indicator and/or data output port
- Built-in over temperature and over current protection with auto-reset capabilities
- Environmental friendliness.

The last point deserves some attention. Many battery types must be recycled or disposed of through some prescribed means. Proper disposal of a battery at the end of its useful life is, thus, an important consideration. Be sure to check the original packaging for disposal instructions. Failure to follow the proper procedures could have serious consequences.

Research has brought about a number of different battery chemistries, each offering distinct advantages. Today's most common and promising rechargeable chemistries include the following:

- **Nickel cadmium** (NiCd). Used for portable radios, cellular phones, video cameras, laptop computers and power tools. NiCds have good load characteristics, are economically priced, and are simple to use.
- **Lithium ion** (Li-Ion). Now commonly available and typically used for video cameras. This battery promises to replace some NiCds for high energy-density applications.
- **Sealed lead acid** (SLA). Used for UPS, video cameras, and other demanding applications where the energy-to-weight ratio is not critical and low battery cost is desirable.
- **Nickel metal hydride** (NiMH). Used for cellular phones, video cameras and laptop computers where high-energy is of importance and cost is secondary.
- **Lithium polymer** (Li-Polymer). When commercially available, this battery will have the highest energy density and lowest self-discharge of common battery types, but its load characteristics will likely only suit low current applications.
- **Reusable alkaline.** Used for light duty applications. Because of its low self-discharge, this battery is suitable for portable entertainment devices and other non-critical appliances that are used occasionally.

No single battery offers all the answers; rather, each chemistry is based on a number of compromises.

A battery, of course, is only as good as its charger. Common attributes for the current generation of charging systems include quick-charge capability and automatic battery condition analysis and subsequent *intelligent* charging.

KEY TERMS

The following terms are commonly used to specify and characterize batteries:

Energy Density. The storage capacity of a battery measured in *watt-hours per kilogram* (Wh/kg).

Cycle Life. The typical number of charge-discharge cycles for a given battery before the capacity decreases from the nominal 100% to approximately 80%, depending upon the application.

Fast Charge Time. The time required to fully charge an empty battery.

Self-discharge. The discharge rate when the battery is not in use.

Cell Voltage. The output voltage of the basic battery element. The cell voltage multiplied by the number of cells provides the battery terminal voltage.

Load Current. The maximum recommended current the battery can provide.

Current Rate. The *C*-rate is a unit by which charge and discharge times are scaled. If discharged at 1*C*, a 100 Ah battery provides a current of 100 A; if discharged at 0.5*C*, the available current is 50 A.

Exercise Requirement. This parameter indicates the frequency that the battery needs to be exercised to achieve maximum service life.

Sealed Lead-Acid Battery

The lead-acid battery is a commonly used chemistry. The *flooded* version is found in automobiles and large UPS battery banks. Most smaller, portable systems use the *sealed* version, also referred to as *gelcell* or slA.

The lead-acid chemistry is commonly used when high power is required, weight is not critical, and cost must be kept low. The typical current range of a medium-sized SLA device is 2 Ah to 50 Ah. Because of its minimal maintenance requirements and predictable storage characteristics, the SLA has found wide acceptance in the UPS industry, especially for *point-of-application* systems.

The SLA is not subject to memory. No harm is done by leaving the battery on float charge for a prolonged time. On the negative side, the SLA does not lend itself well to fast charging. Typical charge times are 8 to 16 hours. The SLA must always be stored in a charged state because a discharged SLA will sulphate (a chemical process that ruins the battery). If left discharged, a recharge may be difficult or even impossible.

Unlike the common NiCd, the SLA prefers a shallow discharge. A full discharge reduces the number of times the battery can be recharged, similar to a mechanical device that wears down when placed under stress. In fact, each discharge-charge cycle reduces (slightly)

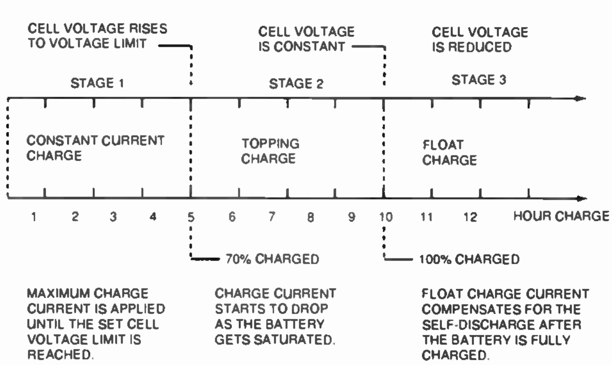


Figure 8.4-10. The charge states of an SLA battery.[3] (*Used with permission.*)

Table 8.4-1
Recommended Charge Voltage Limit for the SLA Battery.[3]

	2.30 V to 2.35 V/cell	2.40 V to 2.45 V/cell
Advantage	Maximum service life; battery remains cool on charge; battery may be charged at ambient temperature exceeding 30° (86°F).	Faster charge times; higher and more consistent capacity readings; less subject to damage because of under charge condition.
Disadvantage	Slow charge time; capacity readings may be low and inconsistent. Produces under charge condition that may cause sulphation and capacity loss if the battery is not periodically cycled.	Battery life may be reduced because of elevated battery temperature while charging. A hot battery may fail to reach the cell voltage limit, causing harmful over charge.

the storage capacity of the battery. This wear down characteristic also applies to other chemistries, including the NiMH.

The charge algorithm of the SLA differs from that of other batteries in that a *voltage-limit* rather than *current-limit* is used. Typically, a multistage charger applies three charge stages consisting of a *constant-current charge, topping charge* and *float charge* (see Figure 8.4-10). During the constant-current stage, the battery charges to 70% in about five hours; the remaining 30% is completed by the topping charge. The slow topping charge, lasting another five hours, is essential for the performance of the battery. If not provided, the SLA eventually loses the ability to accept a full charge and the storage capacity of the battery is reduced. The third stage is the float charge that compensates for self-discharge after the battery has been fully charged.

During the *constant-current charge*, the SLA battery is charged at a high current, limited by the charger itself. After the voltage limit is reached, the topping charge begins and the current starts to gradually decrease. Full charge is reached when the current drops to a preset level or reaches a low-end plateau.

The proper setting of the cell voltage limit is critical and is related to the conditions under which the battery is charged. A typical voltage limit range is from 2.30 V to 2.45 V. If a slow charge is acceptable, or if the room temperature may exceed 30°C (86°F), the recommended voltage limit is 2.35 V/cell. If a faster charge is required and the room temperature remains below 30°C, 2.40 or 2.45 V/cell may be used. Table 8.4-1 compares the advantages and disadvantages of the different voltage settings.

REFERENCES

1. Lawrie, Robert, *Electrical Systems for Computer Installations*, McGraw-Hill, New York, 1988.
2. DeDad, John A., "Auxiliary Power," in *Practical Guide to Power Distribution for Information Technology Equipment*, Intertec Publishing, Overland Park, Kan., pp. 31–39, 1997.

3. Buchmann, Isidor, "Batteries," in *The Electronics Handbook,* J. C. Whitaker ed., p. 1058, CRC Press, Boca Raton, FL, 1996.

BIBLIOGRAPHY

Angevine, Eric, "Controlling Generator and UPS Noise," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, March 1989.

Baietto, Ron, "How to Calculate the Proper Size of UPS Devices," *Microservice Management,* Intertec Publishing, Overland Park, KS, March 1989.

Federal Information Processing Standards Publication No. 94, *Guideline on Electrical Power for ADP Installations,* U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1983.

Highnote, Ronnie L., *The IFM Handbook of Practical Energy Management,* Institute for Management, Old Saybrook, Conn., 1979.

Smith, Morgan, "Planning for Standby AC Power," *Broadcast Engineering,* Intertec Publishing, Overland Park, KS, March 1989.

Section 9: Technical Management and Safety

Introduction by Andy Butler, PBS, Alexandria, VA

Is broadcast engineering truly a unique art? Are there specific skills and a definable knowledge base that set a broadcast engineer apart from traditional single discipline practitioners such as electrical or mechanical engineers? That question is certain to spark lively debate any time a group of engineers gathers. This final section may not settle a bar room debate but it is an excellent resource that fully explores the wide variety of diverse disciplines that broadcast engineers must understand.

Veteran engineering director Marvin Born opens with a realistic look at the business of managing a technical facility. As he observes, "the number one priority of any manager is to insure the financial viability of the company." Born details the diverse issues that successful engineering managers must address to meet this responsibility.

One key to an efficient technical facility is good documentation. Fred Baumgartner and Terry Baun have learned this lesson firsthand. They believe that "the cost of neglecting hours of engineering documentation is paid in weeks of reconstruction." This chapter offers excellent advice on what information to keep and how to keep it.

Broadcast stations are complex facilities housing hundreds of discrete devices that must work together smoothly to produce and deliver a final product. The glue that binds the parts together is skillful systems engineering. Gene DeSantis is a veteran designer who offers some great techniques to make certain the broadcast factory you design delivers the best product for the least possible cost and effort.

Broadcast facilities utilize a number of technologies that are dangerous if not used properly. Transmitters of all types produce non-ionizing radiation that can cause severe injury. Kelly Williams shares his considerable knowledge of the hazard and its management. Finally, Cliff Ferris reviews basic "good practices" to avoid electrical shock. Linda Kelly concludes our safety review with a look at tower safety, fall arrest systems and OSHA rules and regulations.

No matter how well a system is planned and constructed disaster will eventually strike. The key to success is to include risk analysis and strategic recovery planning as a key part of your overall facility planning. Richard Rudman offers a carefully considered game plan for attacking the risk/recovery planning analysis.

This is a rich and exciting section that chronicles many of the unique challenges that broadcast engineers face every day.

9.1
MANAGING A TECHNICAL FACILITY

MARVIN BORN
DISPATCH BROADCAST GROUP, COLUMBUS, OH

INTRODUCTION

There is a standing order in any business that forms the foundation for any manager and that order certainly applies in broadcasting. The number one priority of any manager is to insure the financial viability of the company. This is number one because without the company, no other order, policy or direction is needed. The purpose of this chapter is to enhance the business aspect of the engineering department and how it relates to daily operations of the station, from an accounting-finance point of view rather than from a purely a technical point of view.

ACCOUNTING FUNCTIONS

Control is a word heard in the accounting department of every business. The business office's job is to manage the flow of funds in and out of any business and to keep records of all these movements. One such record is the *general ledger*, which is the master record of what happens in any given month. This is an ongoing month-to-month and year-to-year history of the business. Every expenditure, every utility payment and every hour of labor cost will be accounted for in the general ledger. Various parts of the general ledger can be distributed to the department heads to help in maintaining compliance with their budgets. This information is typically called *audit trails* and they provide detailed information on each and every purchase charged to the engineering accounts. The primary purpose of these audit trails is to match purchase orders with the invoices for the vendors. The audit trails also allow engineering managers to keep a running total of what is actually charged to their accounts in a given month. Due to billing cycles in both the vendors and the business office, it is possible that some invoices will be coded and paid in the expected month. Therefore, in one month an engineering department could be under budget, yet the next month be over budget and have to explain what happened. The audit trails will warn engineering managers of missing invoice payments and they can expect the over budget amount the next month. With this information engineering managers can still spend the normal budgets amount for each month without constantly under spending then over spending to catch up with late billing information.

Purchase Orders

Of all the paperwork engineers are expected to do, *purchase orders* (PO) are one of the least popular, but most helpful. POs are an accounting department method of control that shows that the proper authority to purchase has been obtained. POs provide detailed records of departmental spending and aid in projecting departmental budget needs for the next fiscal year. The engineering manager or their assistant must have the discipline to enter each and every purchase order into a computer database, similar to Figure 9.1-1.

The first column is the vendor name, column B is the date the PO was written and column C is the PO number. Column E is the amount for which the PO was written. These four columns are filled in for each PO when it is written. Set the spreadsheet to total the Column E which will yield a running total for year-to-date.

Columns F, G and H have data entered when the PO is paid—do not skip this entry. It is the most useful section of the database. Column D is the general ledger account code number for the item which can be entered either when the PO is written or at payment. The cellular bill is coded to account 350, which is a general small parts account. Code number 355 is a physical plant account code, etc. Column F is the actually amount charged to the account code and paid to the vendor. There can be some differences between what is written and what is actually paid, such as local taxes being added, etc. These two columns will be an automatic tip-off if more material was ordered than authorized so it also acts as an internal control. This is much faster than physically comparing the PO to the invoice. Column F also provides a variance amount between actual and budgeted figures. Column G is the date of approval for payment. This is handy when bills get lost. It also provides a means to monitor turnaround from purchases to payment. (Some companies allow a discount of one or two percent for early payment.)

The final column is a description of the product, such as small parts, heads, cleaning material. Typically the department's annual operating budget has the account codes broken down into individual items, that have high dollar value such as video heads, audio heads, transmitter tubes, etc. If this description column in the PO database matches the operating budget items, the database can be sorted by items.

As a matter of fact this is the whole basis of this extensive database. Sort the database by account code, October PO's

A	B	C	D	E	F	G	H
Vendor	Date	PO Number	Code	Amount	Actual Paid	Date paid	Description
Abekas	10/21/97	5568	350	175			sm parts
Airtouch Cellular	10/7/97	5834	350	47.59			sm parts
Alltel	10/7/97	5518	350	77.32			sm parts
Alltel	10/8/97	5531	355	514.26			sm parts
Ameritech	10/15/97	5865	399	450			installation
Audio Encounters	10/1/97	5485	350	157.57	157.57	10/10/97	sm parts
Audio Image	10/1/97	5237	350	390	450.89	10/10/97	sm parts
Audio Implements	10/21/97	5562	350	280			sm parts
Black Box	10/21/97	5559	350	93.35			sm equip
Buckeye Truck	10/1/97	5824	385	30	19.02	10/10/97	service
Burns Security	10/17/97	5557	350	2000			secruity
Carl Zipf	10/1/97	5822	355	23.8			building
Central Communications	10/15/97	5837	350	676.81			service
Columbus Fasteners	10/3/97	5523	350	42.44	42.44	10/10/97	ship supp
Columbus Paper	10/15/97	5828	365	190			office supp
Running totals				$5,148	669.92		

Figure 9.1-1. Sample purchase order database.

then by description item. Separate the account codes with several spaces and insert totals for each code. This yields the exact amount spent on each code and each item the previous year. Since broadcast operations do not change that much from year to year, the major part of figuring the operating budget is finished. Additional columns can be added for the number of anticipated future items.

As the years pass, the department will have a number of previous databases to use to project spending needs. Try stripping out the amount paid and the item description and make a new database with the description in column A and each of the years in Column B, C, D, etc. By graphing the amount paid, the department can project the present increase each vendor has been taking each year. This data will help the department defend its budget at the annual meetings.

CAPITAL

The purchase of capital for a broadcast facility is one of the most important functions of a chief engineer and his associates. Not only does capital purchasing have the most of the department's resources at stake, it determines future operating and competing ability for several years to come. A typical method of equipment purchase involves the due diligence study of all specifications, performance history and technical acceptability of equipment. This section will skip the technical methods of equipment selection and move to a more financial method of selecting equipment. The purpose of this section is to remove the emotional factor in equipment selection. It also removes the language barrier between engineers' methodology and the general managers' sales language orientation.

The financial method of evaluation involves analyzing the flow of cash into and out of a piece of equipment. There are three methods of determining the cash benefits of a capital purchase. They are:

- Payback
- Net present value
- Internal rate of return

Payback is the most common and easiest understood method. Simply stated, payback is the amount of time required to recover the capital outlay used to make a purchase. For example, a station purchases a new edit suite, installs it and determines the amount of money that system can generate per session. Assuming one session per day, the payback period is the number of days, months, etc. that pass before the initial cash outlay is returned. The longer the period of time, the less desirable the purchase. Payback does not account for the time value of money. In other words, any interest the money would have generated during the payback time is not accounted for. If the interest were included the period would have been shorter. Further, secondary cash out flows are not included, such as maintenance costs and tax benefits or costs, the most important of which is depreciation. While payback is included here for discussion and educational purposes it is not considered a reliable method of financial analysis due to its short comings.

Net present value (NPV) is a more advanced method of evaluating the performance of a capital investment. Unlike payback, NPV accounts for the cash flows,

SAMPLE NET PRESENT VALUE TABLE

Flow		BT	Tax	AT	Time	Factor	PCFCF	Direction
Buy new		100,000	1	100,000	0		100,000	out
Sell old		10,000	1	10000	0		10000	in
		COST					90,000	
Revenue								
	new	80,000	0.6	48,000	1 to 5	3.79079	181957	in
	old	40,000	0.6	24,000	1 to 5	3.79079	90978	out
Expense								
	new	10,000	0.6	6,000	1 to 5	3.79079	22744	out
	old	8,000	0.6	4,800	1 to 5	3.79079	18195	in
Depreciation								
	new	15,000	0.4	6000	1	0.90909	5454	in
		22,000	0.4	8800	2	0.82645	7272	in
		21,000	0.4	8200	3	0.75131	6160	in
		21,000	0.4	8200	4	0.68301	5600	in
		21,000	0.4	8200	5	0.62092	5091	in
	old	6000	0.4	2400	1 to 5	3.790779	9097	out
	PRESENT VALUE FUTURE CASH FLOWS						106910	
	NET PRESENT VALUE (PVFCF minus COST)						16910	

Figure 9.1-2. Sample net present value table.

over time, both into and out of the financial analysis of the investment. These cash flows are then evaluated in terms of the time value of money.

For example, assume a station is buying a new transmitter for $100,000. It will sell the old transmitter to a local radio station for $10,000. Further it expects this new transmitter to have an expected useful life of five years. It pays 10% interest at the bank. It also plans to sell the new transmitter at the end of five years for $15,000. (Note, five years and 10% cost of capital are used as standard values. refer to the NPV chart Figure 9.1-2).

Flow indicates what the money is doing, BT is a column for values before tax and AT is after tax. Tax is either 1 for no tax, or .4 for 40% tax rate. .6 is tax of .4 subtracted from one (it is what is left after the tax is paid). *Factor* is the value of one dollar at 10% interest for the number of periods indicated in the time column. The values in the Factor column come from a standard interest table. These tables are available in any financial book or can be calculated by most spread sheet software. The PCFCF column is the value in AT multiplied by the factor value. (The amount of money times the value of one dollar for the period of time.) *Direction* refers to which way the money went IN or OUT.

The *Buy New* column is the $100,000 for the new transmitter. There is no income tax involved with the purchase so tax is "1." *Time 0* is the present and there is no time value of money factor for this cash flow. *Present Value Future Cash Flow* (PVFCF) is the actual dollar value of the cash flow corrected for any tax and computed for its value after a period of time. Time 1 is one year, time 2 is two years etc. Time 0 is the present time. There is no tax and no interest due, therefore the present value of the cash flow is $100,000. The "out" in the last column shows that the cash was flowing out. Money was paid (out) for the transmitter.

The second line shows that the old transmitter was sold for $10,000. The direction column shows IN; therefore $10,000 flowed into the analysis and the bank. The $10,000 is subtracted from the new cost for a total of $90,000 spent on the transmitter.

The revenue section is a little more difficult to understand. Using an example, assume the following. The new transmitter will have more coverage than the old one; therefore, it will generate $80,000 per year for each year of service. Since the tax rate is 40%, multiply the $80,000 by .6 (100% less 40% equals 60%, which is equal to .6) which yields $48,000 actual income per year. Multiply the income by the factor for five years and a total of $181,957 which is generated by the new transmitter. Since this is income the direction is IN.

Next, allow that the old transmitter was generating $40,000 per year, times the tax for an income of $24,000. Multiply 24,000 by the factor for a value of $90,978. However, since the station did not really get this money, it is cash flow OUT. For correct analysis the money that the old transmitter earned must be considered.

The expenses involved with the new transmitter and the expenses that would have incurred by the old transmitter must also be considered. Since the new transmitter has more power, it can be assumed that it requires larger tubes at a cost of $10,000 per year. The old transmitter was lower power and its cost of maintenance was only $8,000.

Taking into consideration the taxes on both cash values, then the factor for five years, the $10,000 becomes $22,744 and the $8,000 becomes $18,195. Most important is the direction for the $10,000 is OUT since we spent the money. The $8,000 is IN, money was not spent. Simply put, expenses are money out, whereas revenue is money in.

To demonstrate depreciation, consider a five year method for the new transmitter, but simple depreciation for the old transmitter. Depreciation is a tax benefit that the tax code allows businesses to regain the cost of new equipment through reduced taxes.

The five year method allows a value to be assigned for each year. In this case $15,000 the first year and $22,000 for year two and $21,000 for years three, four and five. The total will be the cost of the new transmitter. Each of the yearly values will be computed for the after tax amount, multiplied by the correct factor yielding the present value in each of the years. Two things are noteworthy here. First, since this is a tax value being returned, the actual tax paid must be computed (not what is left after tax); therefore use .4 as the tax rate which is equivalent to 40%. The factor is from the time value of money tables for one dollar per year. In this case one dollar in year one will be .90909 cents. The direction is IN since cash is received into the analysis. Following the first depreciation line, the tax on $15,000 was $6,000. Project that value to one year in the future, the $6,000 has an actual PCFCF of $5,454. Each of the years two through five are computed the same way.

The old transmitter will be depreciated via the older straight line method. The tax value was $6,000 which represents $9,097 after five years. Notice this was a cash out, since it was money the station will not get, since it sold the transmitter.

Now the whole analysis has been completed. The present value of all future cash flows are $106,910. In simple terms, the new transmitter will generate $106,910 in revenue and depreciation less expenses. However, the actual cost of the new transmitter, $90,000, must be subtracted. The actual money into the business will be $16,910.

The term applied to this final value is Net Present Value (NPV). This simulation of buying a new transmitter has a positive NPV and would be a good business decision financially. The larger the NPV (if it is positive) the better the expenditure is for the company. When the NPV is negative, the company will actually lose money on the deal and would require serious consideration before proceeding with the purchase. As the NPV approaches zero, the capital purchase makes no financial difference.

NPV is a valuable tool to help with the decision making process in purchasing capital equipment. NPV calculations are not the normal domain of the broadcast engineer. Chief Engineers tend to think in terms of how a purchase will enhance the signal, have more coverage or allow expanded production. Their typical presentation at the annual capital meeting is based on technical parameters and not business methods. If the engineer can also provide a sound revenue based analysis along with the technical analysis, he will have a much stronger base for his arguments for purchases. Since general managers or corporate presidents tend to think in financial terms and not technical terms, they will better understand the NPV analysis. NPV is extremely useful in discouraging emotional purchases.

The purpose of this discussion was to present the availability of alternative tools for engineering management. The example presented was brief and simple. The interest rate and tax rates were chosen to make the math easy—they should not be considered typical or real-world. One should consult a financial management text for complete details or have the station's business manager or corporate treasurer help with the analysis the first time through.

The engineering manager must also present their case for the purchase of new equipment. Capital items that are requested by other departments are easier to present. The purchase of sync generators, distribution amplifiers, waveform monitors and audio consoles can be hard to sell. These items are essential for the operation of the station, but they are hidden and they do not directly generate income or ratings. They are not glamorous like SNG and ENG trucks.

To do the pitch, first prepare a normal engineering presentation. Include all the frequency response and noise information, add ease of maintenance and ease of operation. Outline the highlights of the presentation. Now go through the outline and remove every technical term and replace it with plain language.

Take each of these points and apply them to cash flow. For example, the present audio board causes some slight distortion in the program material and has a bad noise figure by today's standards. Point out that this noise and distortion can cause young (age 18 to 34) women to tune away from the station. Then point out how a one and five percent increase in that demographic would affect the ratings and how the ratings increase could affect the revenue. Engineers do not normally work in this area; however, the general manager does not know what a lack of frequency response or system noise can do to the listening audience. If you present the new audio board as having low noise and a frequency response of dc to light, that GM will not connect his low female ratings with the problems in the audio board. While this is a simplistic example, it serves the point. Speak the language of business and present a pitch in that language. Relate how your desired product will either increase revenue or decrease expenses.

THE BUSINESS OF LIVE SHOTS

The Department of Transportation (DOT) imposes a number of rules regarding the driving of trucks requiring a Commercial Drivers License (CDL). While all of the rules apply to the SNG truck driver there are a couple of quirks that make complying with the law difficult. A commercial, over-the-road driver's primary duty is to drive the truck. He can drive for 8 hours and stop to rest and make the appropriate entries in the truck log. The SNG driver's primary duty is not the driving of the truck. It is the uplinking of television information. The truck is a means to that end. The typical scenario is to drive a couple of hundred miles to a news story, set up the truck, handle the story and return home. What causes problems here is the working, but not driving time. A driver can only work fifteen hours before requiring eight hours rest. That is fifteen clock hours not just behind the wheel. For example, the driver arrives at work at 5:00 am and spends an hour preparing his vehicle and then drives five hours to the site of a football game arriving at 11:00 am. He then spends the next four hours setting up the truck and installing cables. He then runs the game which ends at 5:30 pm. His news department requests that he stays for a 6:20 sports live shot, he then spends two hours cleaning up and packing the truck. It is now 8:30 pm. The driver heads home and arrives at 1:30 am the next morning. The driver had worked twelve and a half hours by the end of the game. The live shot and cleanup made the total of fifteen hours.

According to DOT rules, drivers cannot drive a truck after 15 hours without eight hours of rest. This driver should never get behind the wheel let alone drive five hours. In this case the driver was timed out at the end of the game. The manager of an SNG or remote truck that required CDL licenses has to plan ahead and assign a second person to drive the return trip home or allow the driver to stay over. Having a second driver riding in the truck does not count as rest unless the truck has a sleeper berth. In other words, the second driver must be off duty (away from the work site and the game) for a full eight hours before driving the truck or at least have enough driving time left to make it all the way home.

A second more subtle way to time out is the "70 hours in ten days rule." If a driver works 70 hours in a ten day period the driver must have a 24-hour rest period. For example, an SNG driver works a normal five day week and clocks 40 hours. The driver goes home on Friday, but is called out on Saturday, for an SNG remote out of the city. The driver drives all day Saturday, goes on a Sunday remote and has a good night's sleep. The driver is then sent on to another and another remote which together lasts the rest of the week. Each night the driver gets the minimum eight hours rest. The driver arrives back home the following Sunday. What rule was violated? The driver worked more than 70 hours in a 10 day period and drove the truck without the required 24 hours off.

RF

Our CDL scenario concerned the truck driving only. Common sense would indicate that tired operators are more prone to mistakes and the RF equipment on board an SNG truck can provide a real danger. The output of a C- or Ku-Band uplink can easily exceed the maximum RF levels allowed by the FCC.

When a truck is purchased, have it measured for RF levels all around at ground level and on top of the truck. Each employee should carry an active RF monitor with an audible alarm. This device will sound if the employee enters an RF field that exceeds the maximum level. As a minimum, do not allow anyone on the truck when the transmitter is active or, even better, do not buy a SNG truck with a roof platform.

The voltages that operate the klystron amplifiers or traveling tube amplifiers are lethal. Every precaution that is applied to a broadcast transmitter applies to the SNG. Further, trucks are more dangerous, the space is cramped, there is a pressure situation and the operator/driver will have been working longer hours. This is a dangerous combination.

As an engineering manager one should try to anticipate what other people will do and plan for a safe way to service that anticipated need. The typical SNG is not designed to carry passengers in the operating area. However, in a pressure news story someone will suggest they edit while the truck is being driven to the next location. While the danger may be obvious to an engineering manager, other departments may not understand and insist that they be allowed to work in the truck while it is moving. Cover this with the news director and every truck driver, making the driver responsible for the truck safety as well as its operation. Make sure that they have the support to say no to unsafe operations regardless of the believed emergency presented by people who want to take a shortcut.

BUSINESS MANAGEMENT

Somewhere it is said that of the three areas of business administration, accountants record the past, managers handle the present and finance people plan the future. All three aspects are important in their own right and a good engineering manager will know how to use each. The typical engineering manager uses them in a different order. Manage today and provide costs to the accounting plan for the future in a technical way.

Budgeting

There are three documents that relate to accounting: the income statement, the balance sheet and the cash flow statement. Normally, these documents are not available to the engineering department as a whole. A good engineering manager will know of these documents and how they are constructed. They will also know how operations within their department affect the line items on these documents.

The purpose of a balance sheet (shown in Figure 9.1-3) is to account for each dollar within the company. The left side contains all the assets of the company (cash in) such as cash, accounts receivable, property and equipment inventory and of course stockholder equity. While there are many items that can be placed in the asset side, those are the major ones. On the other side are the liabilities, money the company is expected to pay (cash out). The major items will be accounts payable, debt, (short term and long term), accrued taxes and dividends payable. When all the

TELEVISION BUILDERS INC.
Balance Sheet, December 31, 1996

Cash	$11,000	Accounts payable	$18,800
Temporary investments	8,000	Accrued wages payable	10,550
Prepaid expenses	1,200	Income taxes payable	4,650
Accounts receivable	25,500	Mortgage payable	50,000
Inventory	36,300	Common stock	120,000
Plant assets	168,000	Retained earnings	46,000
Total Assets	250,000	Total Equities	$250,000

Figure 9.1-3. Sample balance sheet.

TELEVISION BUILDERS INC.
Balance Sheet, December 31, 1996

Cash	$12,000	Accounts payable	$19,800
Temporary investments	8,000	Accrued wages payable	10,550
Prepaid expenses	1,200	Income taxes payable	4,650
Accounts receivable	25,500	Mortgage payable	50,000
Inventory	36,300	Common stock	120,000
Plant assets	168,000	Retained earnings	46,000
Total Assets	251,000	Total Equities	$251,000

Figure 9.1-4. Sample balance sheet.

accounting is finished, assets should equal liabilities. While the engineering department does not normally see a balance sheet, the engineering department can affect the areas of cash, expenses and inventory as well as plant assets.

If the engineering manager is ordering a large capital purchase that requires a down payment, a note should be sent to accounting informing them of the capital (cash) need in advance. Good business managers will manage the cash such that no large amounts are present in non-interest bearing accounts. They will need some time to transfer funds to the proper account for the down payment. This is the purpose of the cash flow statements. A quick conversation with the business manager will tell how much time is needed for the cash management functions.

Businesses are cyclic, for example, fall has football coverage plus a major rating period and managers expect their staff and maintenance requirements to rise. The same cyclic pattern affects the business office, however, it is typically 30 to 60 days behind. In other words, the business office is paying the bills from the previous two months. While engineering related matters are the engineering manager's foremost concern, be aware of the cycle in the business office and get expenses in on time. Holding invoices too long will cause them to be in the next month's accounts payable and affect the balance sheet as shown in Figure 9.1-4.

The income statement is a financial report with only of one column of information. It lists most of the major items previously mentioned; however, the net income line, the infamous bottom line indicates whether a profit was made or lost for the period covered by the statement. Engineering can have a major effect on this bottom line, especially the operating expense line. That line can be affected by spending money but also by saving money with prudent purchasing practices. Indirectly, the engineering department can affect the income line, with tower rental, remote truck rental and support of news and remote operations. Figure 9.1-5 is a typical income statement for a non-broadcast, manufacturing business. It is used because the broadcast example is different and a basic understanding of a typical balance sheet is necessary.

In any business there are two types of costs or expenses—variable costs and fixed costs. In the example balance sheet there is a cost of goods sold, which includes inventory, purchases for parts and the costs of making the product. What is important to remember is that as the number of products made increases, so does the cost of material and labor to make the product. Likewise, if orders for the product fall so do the cost of goods sold. The line item, *Gross profit on sales* is

Income Statement
January 1, 1996–December 31, 1997

Sales	$460,000
Costs of Goods Sold	360,000
Gross profit on sales	100,000
Operating expenses	50,000
Earnings before Taxes	50,000
Interest expense	30,000
Income Before Taxes	20,000
Taxes	8,000
Net Income	12,000
Common Stock Dividend	10,000
Contribution to Retained Earnings	2,000

Typical income statement for a manufacturing business.

	Profitable	Non-Profitable
Gross Sales	460,000	230,000
Sales commission and variable costs	46,000	23,000
Gross Profit on Sales	414,000	207,000
Operating expenses	276,000	276,000
Earning before tax	138,000	−69,000
Interst expense		
Tax		
Net Income (bottom line)		
Stock Dividend		
Contribution to Retained Earnings		

Typical income statement for broadcast station both profitable and taking a loss.

Figure 9.1-5

the result of subtracting the cost of goods sold from the income generated by the selling of the product.

The other type of cost of doing business is fixed costs. In the manufacturing business this is the operating expenses and mortgage interest expense. Operating expenses include the utilities, maintenance of plant and equipment and labor for non-line item employees, such as the business office and management. When the fixed cost is subtracted from the gross profit from sales, it yields the line item marked *income before taxes.* When the correct amount of tax is then subtracted, the result is the net income or bottom line.

What are the differences for broadcasting? Broadcasters do not manufacture a product, they sell time. But that time must be filled regardless of whether that time is a paying commercial message or a non-revenue public service announcement or station promotional announcement. Stated in a different more business manner: broadcast has very low variable costs (usually only sales commissions) but high fixed costs. Broadcasters pay the electric bill and maintain the machines and antenna towers regardless of what passes through the equipment. Visualize the manufacturing income statement presented and lower the cost of goods sold to $46,000 (which for this example represents the commissions to the sales department). The broadcast station makes a nice profit. However, broadcasters' expenses for operating the station are much higher. Maybe $276,000 for operators, parts, utilities, interest on equipment purchases and maintenance. That leaves a before tax profit of $138,000 (still a nice profit). But, what if there is a downturn in the economy and commercial sales drop to half the normal amount to $230,000 with sales commissions of $23,000 for a gross profit of $207,000? The station now has to fix its schedule with non-revenue spots, but the fixed operating costs of $276,000 remain. The station now has a net loss of $69,000. The high fixed costs made the turn between profit and loss occur very quickly.

A good broadcast engineer will understand how fixed costs affect the bottom line and be cooperative in minimizing them during troubled times. The upside is that during near sold out periods the fixed costs remain the same and station profits will soar just as quickly and dramatically in the other direction. A good engineering manager will know the desires of the general manager in regard to driving the bottom line. They should understand why it is necessary to cut overtime and forgo capital equipment for a given year when times are lean and at the same time understand the requirement to drive the net income up to increase retained earnings on the balance sheet. Good general managers and owners will also understand when to keep the retained earning number high and when to return greater sums of money to the company in the support of larger raises and increased capital purchases. High retained income drives up the stock prices if the company is publicly owned. Insuring the financial viability of the company is the number one job of any manager, general manager, engineering or sales manager. And driving up retained earnings is exactly what they are employed to do. Each department manager does their job in a different way, but the primary job is to have a positive number at the bottom line.

9.2
BROADCAST ENGINEERING DOCUMENTATION

FRED BAUMGARTNER
TCI TECHNOLOGY VENTURES, LITTLETON, CO
TERRENCE M. BAUN
CRITERION BROADCAST SERVICES, MILWAUKEE, WI

INTRODUCTION

Broadcast facilities are designed to have as little downtime as possible. Yet inadequate documentation is a major contributor to the high cost of systems maintenance and the resulting widespread replacement of poorly documented facilities. The cost of neglecting *hours* of engineering documentation is paid in *weeks* of reconstruction.

Documentation is a management function every bit as important as project design, budgeting, planning and quality control, and is often the difference between an efficient and reliable facility and a misadventure. If the broadcast engineer does not feel qualified to attempt documentation of a project, the engineer must at the very least oversee and approve the documentation developed by others.

Within the last few years the need for documentation has increased with the complexity of the broadcast systems. Fortunately, the power of documentation tools has kept pace.

With that in mind, this chapter considers exactly how much documentation is appropriate and how to assemble that documentation. This chapter reviews the current state of good engineering practice in such matters, and outlines the elements of making a business case for the very necessary expense of creating good documentation in broadcast plants.

BASIC CONCEPTS

The first consideration in the documentation process is the *complexity* of the installation. A basic radio facility may require almost no formal documentation, while a large satellite or network broadcast facility may require computerized databases and a full time staff dedicated to documentation updates. Most facilities will fall somewhere in the middle.

A second concern is the need for *flexibility* at the facility. Seldom does a broadcast operation get completely rewired because the cabling wears out or fails. Most often, it is the supporting documentation that has broken down, frustrating the maintainability of the system. Retroactive documentation is physically difficult and emotionally challenging, and seldom generates the level of commitment required to be entirely successful or accurate; hence, a total rebuild is often the preferred solution to documentation failure. Documentation must be considered a hedge against such unnecessary reconstruction.

Finally, consider *efficiency* and speed. Documentation is a prepayment of time. Repairs, rerouting, replacements and reworking all go faster and smoother with proper documentation. If your installation is one in which any downtime or degradation of service is unacceptable, then budgeting sufficient time for the documentation process is critically important.

The Manuals

Equipment manuals are the first line of documentation and deserve our attention and respect. Place those manuals in a centralized location in an order that seems appropriate for your station. Most engineering shops file manuals alphabetically, but some prefer a filing system based on equipment placement. (For example, production studio equipment manuals could be filed together under a *Production Studio* label and might even be physically located in the referenced studio.) Make sure the system is consistent. Few things are as frustrating to a technician as being unable to locate a manual when needing something as simple as a part number or the manufacturer's address.

The Documentation of Conventions

The second essential item of documentation is the statement of *Conventions*. This document contains basic information essential to an understanding of the facility. It should be posted in an obvious location and available to all who maintain the plant. Consider the following examples of conventions:

- Where are the equipment manuals and how are they organized?
- What is the architecture of the ground system? Where is the central station ground and is it a star, grid, or other distribution pattern? Are there separate technical and power grounds? Are audio shield grounded at the source, termination or both locations?
- What is the standard input/output architecture? Is this a +8 dBm, +4 dBm, or 0 dBm facility? Is equipment sourced at 600 Ω terminated at its destination, or left unterminated? Are unbalanced audio

sources wired with the shield as ground or is the low side of a balanced pair used for that purpose? How are XLR-type connectors wired—pin 2 high or low?
- How can a technician disconnect utility power to service line voltage wiring within racks? Where are breakers located, and how are they marked? What equipment is on UPS power, generator or utility power? Whom do you call for power and building systems maintenance?
- Where are the keys to the transmitter? Is the site alarmed?

For such an essential information source—it takes very little time to generate the conventions document. Keep it short—a page or two should be sufficient for most installations, and, if located in a obvious place, will keep the technical staff on track and will save service personnel from stumbling around searching for basic information. This document is the key to preventing many avoidable embarrassments.

The conventions document may be framed and posted behind the master control rack, in the engineering shop, or the CE's office. Wherever it is, be certain that it is prominently displayed.

Documentation Systems

The next step beyond the conventions document is a documentation *system.* There are three primary methods: self documentation, database documentation and graphic documentation. In most cases, a mixture of all three is necessary. In addition to documenting the physical plant and its interconnections, each piece of equipment, whether commercially produced or custom made, must be documented in an organized manner. Likewise there are a number of aids for documenting, labeling cables, wiring, etc. that are covered in the following sections.

Self Documentation

In situations where the facility is small and operations are very routine, self documentation is possible. Self documentation relies on a set of standard practices that are repeated. Telephone installations, for example, appear as a mass of perplexing wires. In reality, the same simple circuits are simply repeated in an organized and universal manner. To the initiated, any telephone installation that follows the rules is easy to understand and repair or modify, no matter where or how large—such a system is truly self documenting. Familiarity with telephone installations is particularly useful, because the telephone system was the first massive electronic installation. It is the telephone system that gave broadcasters relay racks, grounding plans, demarcation and virtually all of the other concepts that are part of today's electronic control or communications facility.

The organization, color codes, terminology and layout of telephone systems is recorded in minute detail. Once a technician is familiar with the rules of telephone installations, drawings and written documentation is rarely required for routine expansion and repair. The same is true of many parts of other facilities. Certainly, much of the wiring in any given rack of equipment can be self documenting. For example, a videotape recorder will likely be mounted in a rack with a picture monitor, audio monitor, waveform monitor and vectorscope. The wiring between each of these pieces of equipment is clearly visible, with all wires short and their purpose obvious to any technician familiar with the rules of video. Further, each video cable will conform to the same standards of level or data configuration. Further documentation is largely unnecessary.

By convention, there are rules of grounding, power and signal flow in all engineering facilities. In general, it can be assumed that in most communications facilities, the ground will be a star system, the power will be individual 20 amp feeds to each rack, and the signal flow will be from top to bottom.

Rules that might vary from facility to facility include color coding, connector pin outs, and rules for shield and return grounding. For example, stereo analog audio might be carried on *zip-stereo* cables (two shielded twisted pair wires married together to appear as lamp cord). It would be reasonable that each audio pair uses red for positive, black for negative and the shield is grounded only at the source end. The right channel may be the green pair and the left, the red pair. If this is consistent throughout the facility, only a simple statement of this is necessary to deal with a very large number of interconnections and a document repeating this convention for each circuit would be unnecessary. However, there is no standard for stereo analog audio, and all of the above conventions are reversed, with no ill effect, in many plants.

To be self documenting, the rules must be determined and all of the technicians working on the facility must know and follow the conventions. The larger the number of technicians, or the higher the rate of staff turnover, the more important it is to have a readily available document that clearly covers the conventions in use.

One thing must be very clear: a facility that does not have written documentation is not automatically self documenting—quite the contrary. A written set of conventions and unfaltering adherence to them are the trademarks of a self documenting facility.

But while it is good engineering practice to design all facilities to be as self documenting as possible, realize that there are limits to the power of self documentation. In the practical world, self documentation can greatly reduce the amount of written documentation required, but can seldom replace it entirely.

Database Documentation

As facilities expand in size and complexity, a set of conventions will no longer answer all of the questions. At some point, a wire leaves an equipment rack and its destination is no longer obvious. Likewise, equipment will often require written documentation as to its configuration and purpose, especially if it is utilized in an uncharacteristic way.

Database documentation records the locations of both ends of a given circuit. For this, each cable must be identified individually. There are two common systems for numbering cables—*ascension numbers* and *from-to coding*. In ascension numbering schemes, each wire or cable is numbered in increasing order, one, two, three ... etc. In from-to coding, the number on each cable represents the source location, the destination location, and normally some identification as to purpose and a unique identifier. For example, a cable labeled 31-35-B6 might indicate that cable went from a piece of equipment in rack 31, to another unit in rack 35 and carries black, it is also the sixth cable to follow the same route and carry the same class of signal.

Each method has its benefits. Ascension numbering is easier to assign and commonly available preprinted wire labels can be used. On the other hand, ascension numbers contain no hints as to wire purpose or path, and for that reason *purpose codes* are often added to the markings.

From-to codes can contain a great deal more information without relying on the printed documentation records, but often space does not permit a full delineation on the tag itself. Here again, supplemental information may still be required in a separate document or database.

Whatever numbering system is used, there really must be a complete listing kept in a database of some kind. In smaller installations, this might simply be a spiral notebook that contains a complete list of all cables, their source, destination, demarcations, signal parameters, etc.

Since all cabling can be considered as a transmission line, all cabling involves issues of termination. In some data and analog video applications, it is common for a signal to *loop-through* several pieces of equipment. Breaking or tapping into the signal path often has consequences elsewhere, resulting in unterminated or double terminated lines. While more forgiving, analog audio has similar concerns. Therefore documentation must include information on such termination.

Analog audio and balanced lines used in instrumentation have concerns of their own. It is seldom desirable to ground both ends of a shielded cable. Again, the documentation must reflect which end of a given shield is grounded.

In many cases, signal velocity is such that the length of the lines and the resultant propagation delay is critical. In such circumstances, this is significant information to be retained. In cases where differing signal levels or configurations are used (typical in data and control systems), it is the documentor's obligation to record those circumstances as well.

However or wherever the database documentation is retained, it represents the basic information that defines the facility interconnections and must be available for updating and duplication as required.

Graphic Documentation

Electronics is largely a graphical language. Schematics and flow charts are more understandable than net lists or cable interconnection lists. Drawings, either by hand, drafting machines and tools, or accomplished on computer aided design (CAD) programs are highly useful in conveying overall facilities design quickly and clearly. Normally the wire numbering scheme captured in a CAD document will follow that used in the database documentation, so that the graphic and text documentation can be used together.

CAD drawings are easy to update and reprint. For this reason, documentation via CAD is becoming more popular, even in smaller installations. Since modern CAD programs not only draw but also store drawings, they can effectively serve as an electronic file cabinet for documentation. While there have been attempts to provide electronic/telecommunications engineering documentation templates and corresponding technical graphics packages for CAD programs, most of the work in this area has been done by engineers working independently to develop their own systems.

Obviously the enormous scope of electronic equipment and telecommunications systems make it impractical for a standard CAD package to suit every user. Imagine, for example, a situation involving construction and documentation of a sound recording studio. Here the CAD program would need specifications on typical studio recording equipment (size, weights, power requirements), cable and connectors used, wiring types, speaker types and sizes, acoustical and environmental constraints, etc. Now imagine this same program when used to document a broadcast television control room—or a radio news room. The database required would be enormous and updating it would be a full time job in itself.

There is one commercially available program, VidCAD™ that has attempted to address this problem for the telecommunications industry. The program, which is basically an extension and augmentation of the AutoCAD™ program, claims an 11,000-product equipment database, and the ability to add equipment upon user request. In addition the program includes sophisticated report generating and wire routing capabilities. Visio Corporation also offers a CAD package (Visio). Undoubtedly, other CAD programs will offer similar features if the industry demand is sufficient.

Update Procedures

Since documentation is a dynamic tool, as the facility changes, so too must the documentation. It is common for a technician to improve conditions by reworking a circuit or two. Most often this fixes a problem that should be corrected as a maintenance item. But often, it plants a time bomb wherein a future change, based on no or incomplete documentation of the previous work, will cause problems.

It is essential that there be a means of consistently updating the documentation. The most common way of accomplishing this is the mandatory *change sheet.* Whenever a technician makes a change it is reported back to those who keep the documentation. If the changes are extensive, the use of the *red-line* drawing and *edit sheet* come into play. The original drawing and database printout, respectively, are corrected with a red pen. This then is used to update the original documentation.

In some cases, the updating process can be tied to the engineering reports or discrepancy process. Most facilities use some form of *trouble ticket* to track equipment and system performance and to report and track maintenance. This same form may be used to report changes required to the documentation or errors in existing documentation. The anticipated integration of maintenance software with documentation software will prove invaluable for this purpose.

DOCUMENT HANDLING AND NETWORKS

As of this writing, computer networking has reached the desktops of most broadcast facilities. Running a document manager on your network permits documentation to be tied together, and made available to all who require access to it. Installation of a complete documentation system on a company intranet would seem to be a useful option.

Imagine being able to pull up a single circuit and quickly locate a series of drawings and database entries that describe that circuit. If you desire, links direct you to related circuits. You could back out to overview documentation of the whole system or focus down to a specific circuit from any computer with access to the network. Imagine being able to call up a drawing from another one of your facilities across the country to recall how a particular installation was handled there. The file card could be imported and customized to fit the new installation. With PC remote-access software, you or your technicians might be able to instruct an operator to perform a route-around by viewing the station documentation from your home PC. It is clear that this type of hypertext documentation can be a viable means of moving documentation into a paperless environment.

Label Systems

The simplest and cheapest label system is the wire-tie tag label that can be written upon with a pen. For systems planned with wire numbers assigned in advance of construction, tags using wrap around clear plastic protectors can be computer printed, making them more legible, durable and less likely to be ripped off when pulled. Computer printed tags can even be generated as part of a documentation program.

A recent development is the hand-held labeling system which generates professionally printed labels right at the job site, allowing installers to create the cable labels as the connections themselves are being made. These units are also useful for creating equipment designation labels and in-house asset numbering tags.

Perhaps the height of cable labeling is the use of stamped-in or cable printing. Typically, this method applies a printed marking either with ink or by slightly melting (branding) the cable jacket every foot or so. The obvious benefit is that the cable can be identified anywhere along its path. Unfortunately, the costs and time involved make this method impractical except for very large facilities.

Equipment Documentation

Plant documentation does not end when all of the circuit paths in the facility are defined. Each circuit begins and ends at a piece of equipment which can be modified, reconfigured or removed from service. One must keep in mind that someone unfamiliar with the equipment will eventually be asked to return it to service. For this reason, a documentation file for each piece of equipment must be maintained.

Equipment documentation contains these key elements: the equipment manual (or a reference to its location), the modification record, the configuration information, and ideally, the maintenance record.

The equipment manual is the manufacturer's original documentation. As mentioned previously, the manuals must be organized in such a manner that they can be easily located. Typically, the manuals are kept at the site where the equipment is installed if practical. Remember that equipment with two *ends* such as STLs, RPUs, or remote control systems need manuals at *each* location!

Of course, if a piece of equipment is of custom construction, there must be particular attention paid to creating a manual. For this reason, a copy of the key schematics and documentation is often attached directly to the equipment. Often, this built-in manual will be the only documentation to survive over the years.

Many pieces of equipment, over time, will require modification. Typically the modification is recorded in three ways. First, internally to the equipment. A simple note glued into the chassis may be suitable, or a marker pen is used to write on a printed circuit board or other component. Second, the changes may be recorded in the manual, either inside the cover, or on the schematic or relevant pages. If the manual serves several machines, this may not be appropriate. If this is the case, a third option is to keep the modification information in a separate *equipment file.*

Equipment files are typically kept in standard file folders and may be filed with the pertinent equipment manuals. Ideally, the equipment file is started when the equipment is purchased, and should contain purchase date, serial number, all modifications, equipment location and a record of service.

The equipment file is the proper place to keep the configuration information. An increasingly large amount of equipment is microprocessor based or otherwise configurable for a specific mode of operation. Having a record of the machine's default configuration is extremely helpful when a power glitch (or an operator) reconfigures a machine unexpectedly.

Equipment files should also contain repair records. With most equipment, documenting failures, major part replacements, operating time etc., serves a valuable purpose. Nothing is more useful in troubleshooting than a record of previous failures, configuration, modifications, and, of course, a copy of the original manual.

Operator/User Documentation

The documentation that this chapter addressed up until now has been exclusively in the realm of construction and maintenance. There is yet another type of documentation, that designed for system users.

User documentation provides, at its most basic, instructions on how to use a system. While most equipment manufacturers provide reasonably good instruction and operations manuals for their products, when those products are integrated into a system another level of documentation may be required. Complex equipment may require interface components that need to be adjusted from time to time, or various machines may be incompatible in some data transmission modes—all of which is very essential to the proper operation of the system.

Such information often resides only in the heads of certain key users and is passed on by word of mouth. This level of informality can be very dangerous, especially when changes take place in the users or maintenance staff, resulting in differing interpretations between operators and maintenance people about how the system normally operates. Maintenance personnel will then spend considerable time tracking hypothetical errors reported by misinformed users.

A good solution is to have the operators write an operating manual and provide a copy to the maintenance department. Such documentation will go a long way toward improving interdepartmental communications and should result in more efficient maintenance as well.

The True Cost of Documentation

There is no question that documentation is expensive—in many cases it may equal the cost of installation. Still, both installation and documentation expenses pale in comparison to the cost of equipment as well as revenue losses due to system downtime.

Documentation must be seen as a management and personnel issue of the highest order. Any lapse in the documentation updating process can result in disaster. Procrastination and the resulting lack of follow through will destroy any documentation system and ultimately result in plant failures, extended downtime, and premature rebuilding of the facility.

Making a business case for documentation is similar to making any business case. Gather together all the costs in time, hardware, and software on one side of the equation, and balance this against the savings in time and lost revenue on the other side. Like transmission line theory, the most efficient operation occurs when the load matches the source—or in this case, when the cost of documentation is equal to the savings.

Engineering managers are expected to project costs accurately, and the allocation of sufficient resources for documentation and its requisite updating is an essential part of that responsibility.

Documentation Aids

Telecrafter Products, 12687 W. Cedar Dr., Suite 105, Lakewood Co. 80228. 800-257-2448 (plastic tags)

Patch Prints, 4742 San Fernando Rd., Glendale, CA 91204 818-241-5585 (patch bay labeling)

Software/CAD Services

VidCAD Documentation Programs
VDP, Inc.
749 Carver Road
Las Cruces, New Mexico 88005

Visio Corporation
520 Pike Street, Suite 1800
Seattle, WA 92101

InterCAD Company
401 Hudson Road
Sudbury, MA 01776

Teknologica Corp.
PO Box 850021
New Orleans, LA 70185

Portable Label Systems

W. H. Brady Co.
6555 West Good Hope Road
Milwaukee, WI 53223

Brother International Corp.
PO Box 341332
Barlett, TN 38184

9.3
SYSTEMS ENGINEERING CONCEPTS

GENE DESANTIS
DESANTIS ASSOCIATES, LITTLE FALLS, NJ

INTRODUCTION

Modern systems engineering emerged during World War II. Due to the degree of complexity in design, development and deployment, weapons evolved into weapon *systems* during this time. In the 1960s, the challenges of the space program made a systems engineering approach to design and problem solving even more critical. Indeed, the Department of Defense (DoD) and the National Aeronautics and Space Administration (NASA) are two of systems engineering's staunchest practitioners. With the increase in size and complexity of radio and television systems during that same period, the need for a systems approach to planning, designing and building facilities gained increased attention.

Today, large engineering organizations also use a systems engineering process. Much has been written about systems engineering practices in manuals, standards, specifications, and instructional materials. In 1969, the DoD published MIL-STD-499, to assist government and contractor personnel involved in defense acquisition programs. In 1974, this standard was updated to MIL-STD-499A, which specified the application of systems engineering principles to military development programs and is still in use today.

Likewise, the builders of turnkey radio and television systems and facilities have adopted their own unique systems engineering approaches to projects. The tools and techniques of these processes continue to evolve so that each job can be done better, faster and more economically.

This chapter describes systems theory in a general sense, discusses its application to broadcast television and radio systems engineering and concludes with some practical examples and implementations of the process.

SYSTEMS THEORY

In the broadcast industry, systems engineering can be applied to the engineering of audio, video, radio frequency (RF), control, time code, telecommunications, computer systems and software. Building and vehicle systems, including space planning, power and lighting, environmental control and safety systems, can also benefit from the systems engineering approach. All of these systems are made up of interconnected component elements that are programmed to function together in a facility.

For the purpose of this discussion, a *system* is defined as a set of related elements that function together as a single entity. *Systems theory* consists of a body of concepts and methods that guide the description, analysis and design of complex entities.

Decomposition is an essential tool of systems theory. Using an organized methodology, the systems approach breaks down large complex projects into simpler, more manageable component elements, which are treated, analyzed and designed separately. In the end, all of the components are recombined to build the whole.

Holism is an element of systems theory that recognizes that the end product is greater than the sum of its component elements. In systems theory, modeling and analytical methods take into account all essential effects and interactions within a system and between a system and its surroundings. This avoids errors that result when parts of a system are treated in isolation or consideration is focused on a single aspect of the system.

Another holistic aspect of system theory describes *emergent properties*. These are properties that result from the interaction of system components but are not properties of the components themselves.

Though systems theory deals with concrete systems, *abstraction* is an important feature of systems models. Components are described in terms of their function rather than their form. Graphic models such as block diagrams, flow diagrams, timing diagrams and the like are commonly used.

Mathematical models may also be employed. Systems theory shows that, when modeled in abstract formal language, apparently diverse kinds of systems show significant and useful *isomorphisms* of structure and function. Similar interconnection structures occur in different types of systems. Equations describing the behavior of electrical, thermal, fluid and mechanical systems are often essentially identical in form.

Isomorphism of structure and function implies isomorphism of behavior of a system. Different types of systems exhibit similar dynamic behavior such as response to stimulation.

The concept of *hard* and *soft systems* appears in systems theory. In hard systems, mathematical models can describe the components and their interactions. Soft systems cannot be described so easily. They are mostly human activity systems that imply unpredict-

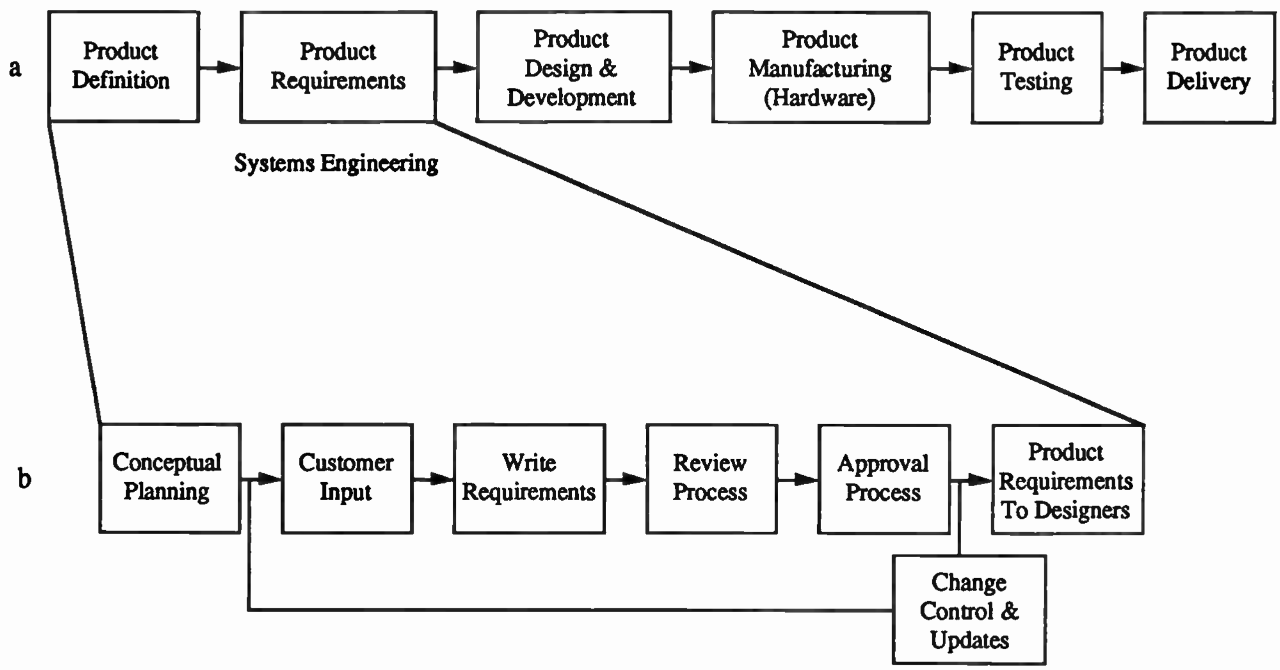


Figure 9.3-1. Product development process (a) showing requirements documentation process (b).

able behavior and non-uniformity. They introduce difficulties and uncertainties of conceptualization, description and measurement. The kinds of system concepts and methodology described above cannot be applied to soft systems.

Systems Engineering

Systems engineering depends on the use of a *process methodology* based on systems theory. To deal with the complexity of large projects, systems theory breaks down the process into logical steps. Even though underlying requirements differ from program to program, a consistent, logical process can be used to accomplish system design tasks. The basic product development process is illustrated in Figure 9.3-1, which shows systems engineering beginning with the description of the product to be designed. Systems engineering encompasses four activities: functional analysis, synthesis, evaluation and decision and a description of system elements (see Figure 9.3-2). The process is iterative, so that with each successive pass, the product element description becomes more detailed. At each stage in the process, a decision is made whether to accept, make changes or return to an earlier stage of the process and produce new documentation. The result of this activity is documentation that describes all system elements and can be used to develop and produce the elements of the system.

Functional Analysis

A systematic approach to systems engineering will include elements of systems theory (see Figure 9.3-3).

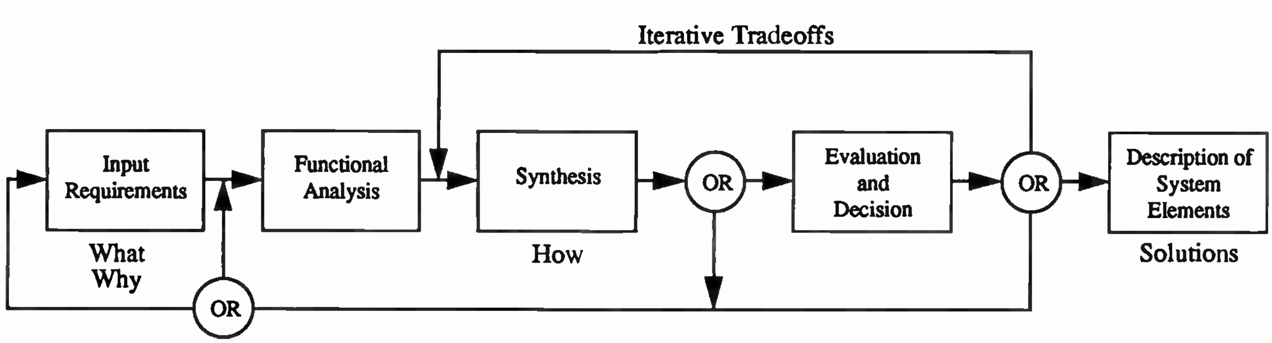


Figure 9.3-2. The systems engineering process. (Adapted from Hoban, F. T., and Lawbaugh, W. M., *Readings in Systems Engineering*, NASA, Washington, DC, pg. 10, 1993.)

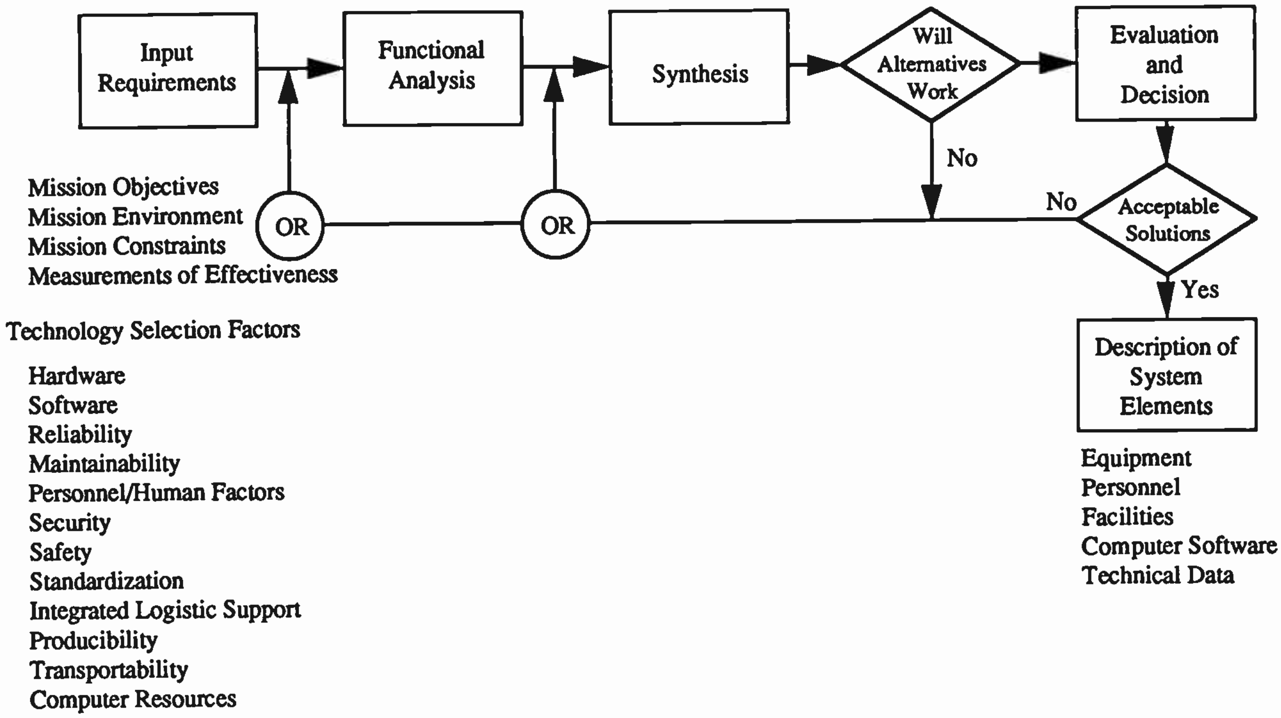


Figure 9.3-3. The systems engineering decision process. (Adapted from Hoban, F. T., and Lawbaugh, W. M., *Readings in Systems Engineering*, NASA, Washington, DC, pg. 12, 1993.)

To design a product, hardware and software engineers need to develop a vision of the product and, ultimately, the product requirements. These requirements are typically based on customer needs researched by a marketing department. (An organized process to identify and validate customer needs will help minimize false starts.) First, system objectives are defined. This may take the form of a mission statement that outlines the objectives, constraints, mission environment and means of measuring mission effectiveness;.

Once the purpose of the system is defined, analysis identifies the requirements and the essential functions the system must perform and why. The functional flow block diagram, a basic tool used to identify functional needs, shows logical sequences and relationships between operational and support functions at the system level. Other functions such as maintenance, testing, logistics support and productivity may also be required in the functional analysis. The functional requirements will be used during the synthesis phase to show the allocation of the functional performance requirements to individual system elements or groups of elements. Following evaluation and decision, the functional requirements provide the functionally oriented data required to describe the system elements.

Analysis of time critical functions is also a part of this functional analysis process, with identification of functions that have to take place sequentially, or concurrently, or on a particular schedule. Timeline documents support the development of requirements for operation, testing and maintenance functions.

Synthesis

Synthesis is the process by which concepts are developed to achieve the functional requirements of a system. Performance requirements and constraints, as defined by the functional analysis, are applied to each individual element of the system, and a design approach is proposed for meeting the requirements. Conceptual schematic arrangements of system elements are developed to meet system requirements. These documents can be used to develop a description of the system elements and can be used during the acquisition phase.

Modeling

Modeling is the starting point of synthesis. Since we must be able to weigh the effects of different design decisions to make choices between alternative concepts, modeling requires determining the quantitative features that describe the operation of the system. We would, of course, like a very detailed model with as much detail as possible describing the system. Reality and time constraints, however, dictate that the simplest possible model be selected to improve our chances of design success. The model itself is always a compromise and is restricted to those aspects that are important in the evaluation of system operation. A model might start off as a simple block diagram with more detail added as needed.

Dynamics

Most system problems are dynamic in nature. The signals change over time, and the components determine the dynamic response of the system. The system behavior depends on the signals at a given instant, as well as on the rates of change of the signals and their past values. The term "signals" can be replaced by substituting human factors, such as the number of users on a computer network, for example.

Optimization

The last concept of synthesis is optimization. Every design project involves making a series of compromises and choices based on relative weighting of the merit of important aspects. Deciding the importance of various features is often subjective but, ideally, the best candidate among several alternatives will be selected.

Evaluation and Decision

Program costs are determined by tradeoffs between operational requirements and engineering design. Throughout the design and development phase, decisions are made based on evaluation of alternatives and their effect on cost. One approach attempts to correlate the characteristics of alternative solutions to the requirements and constraints that make up the selection criteria for a particular element. The rationale for alternative choices in the decision process is documented for review. Mathematical models or computer simulations may be employed to aid in this evaluation decision making process.

Trade Studies

A structured approach is used in the trade study process to guide the selection of alternative configurations and insure that a logical and unbiased choice is made. Throughout development, trade studies are conducted to determine the best configuration that will meet program requirements. In the concept exploration and demonstration phases, trade studies help define the system configuration. They are also used as a detailed

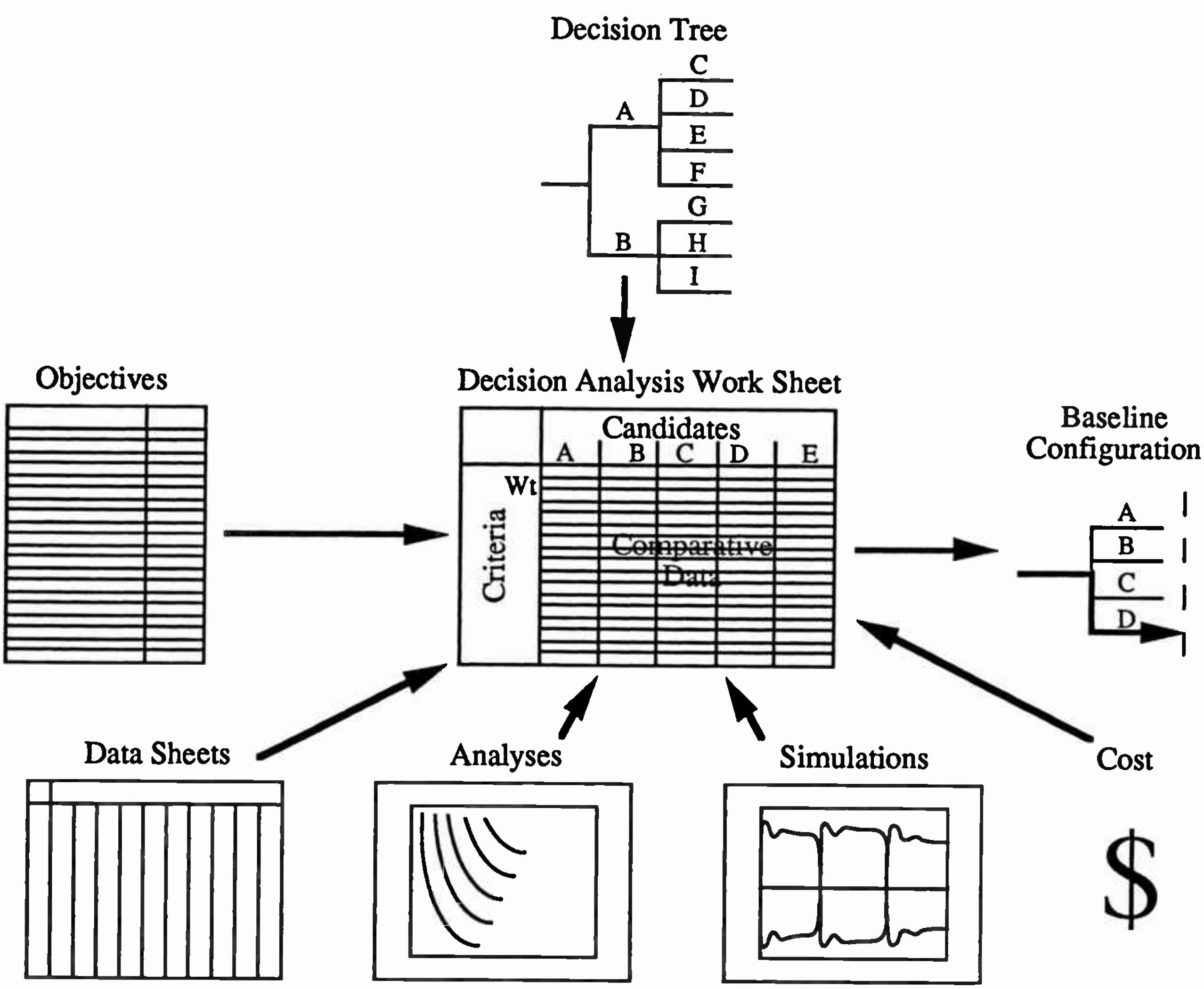


Figure 9.3-4. Trade studies use a systematic approach to decision making. (Adapted from *System Engineering Management Guide*, Contract No. MDA 903-82-C-0339, Defense Systems Management Collage, Virginia, 1983.)

design analysis tool for individual system elements in the full-scale development phase. During production, trade studies are used to select alternatives when changes need to be made. Figure 9.3-4 illustrates the relationship of the various elements that may be employed in a trade study.

Figure 9.3-5 is a flow diagram of the trade study process. To provide a basis for the selection criteria, the trade study's objectives must first be defined. Functional flow diagrams and system block diagrams are used to identify trade study areas that can satisfy certain requirements. Alternative approaches to achieving the defined objectives can then be established.

Complex approaches can be broken down into several simpler areas and a decision tree or *trade tree* constructed to show relationships and the dependencies at each level of the selection process (see Figure 9.3-6). Several trade study areas may be identified as possible candidates for accomplishing a given function, and several alternatives may be candidates for solutions in a given area. The selected candidates are then submitted to a systematic evaluation process intended to weed out unacceptable candidates. Criteria for desirable characteristics are established, although undesirable characteristics may also be included to aid in the evaluation process. Weights are assigned to each criterion to reflect its value or impact on the selection process. This subjective process should also take into account cost, schedule and hardware availability restraints, which may limit the selection.

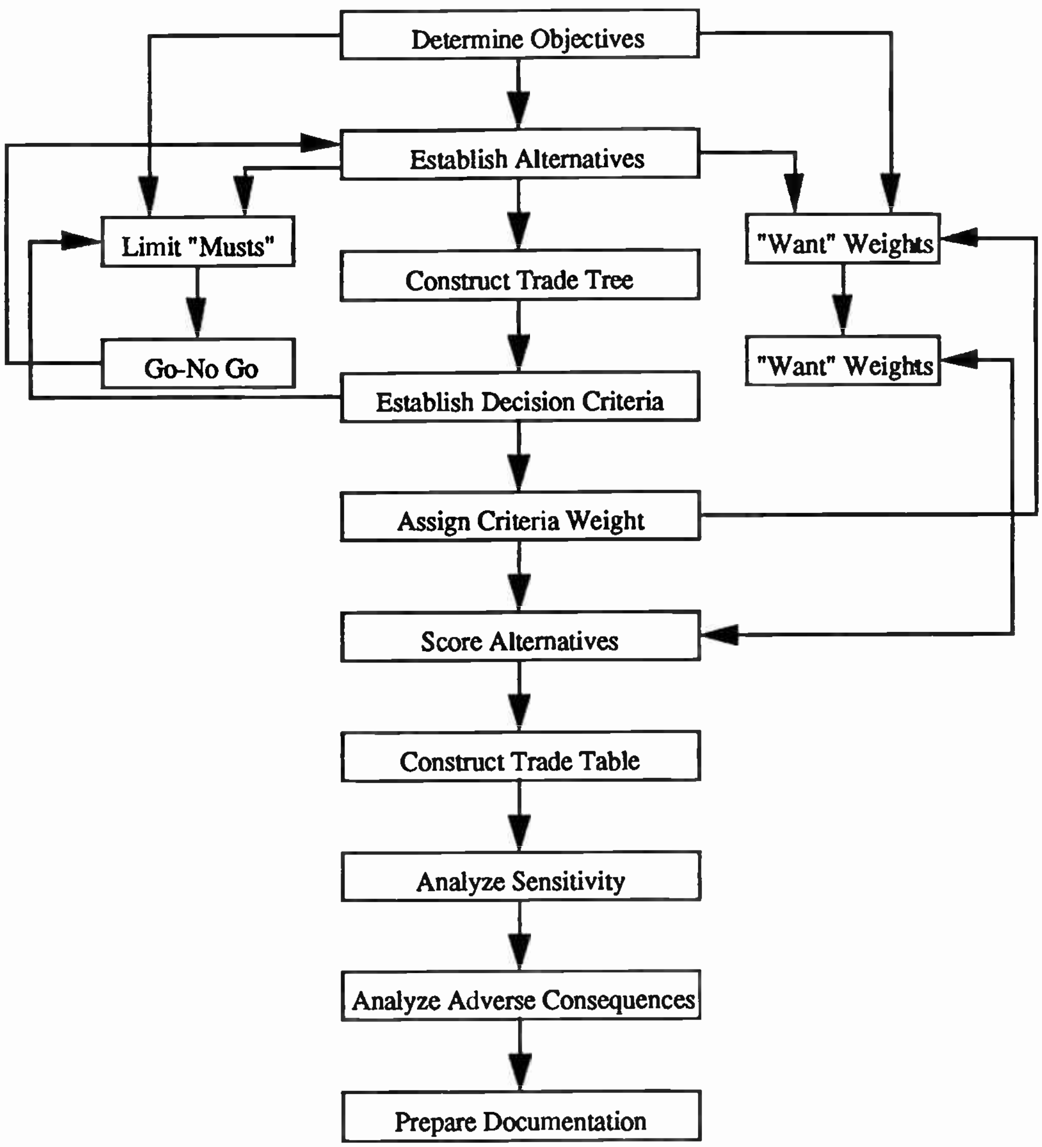


Figure 9.3-5. Trade study process flow. (Adapted from *System Engineering Management Guide*, Contract No. MDA 903-82-C-0339, Defense Systems Management Collage, Virginia, 1983.)

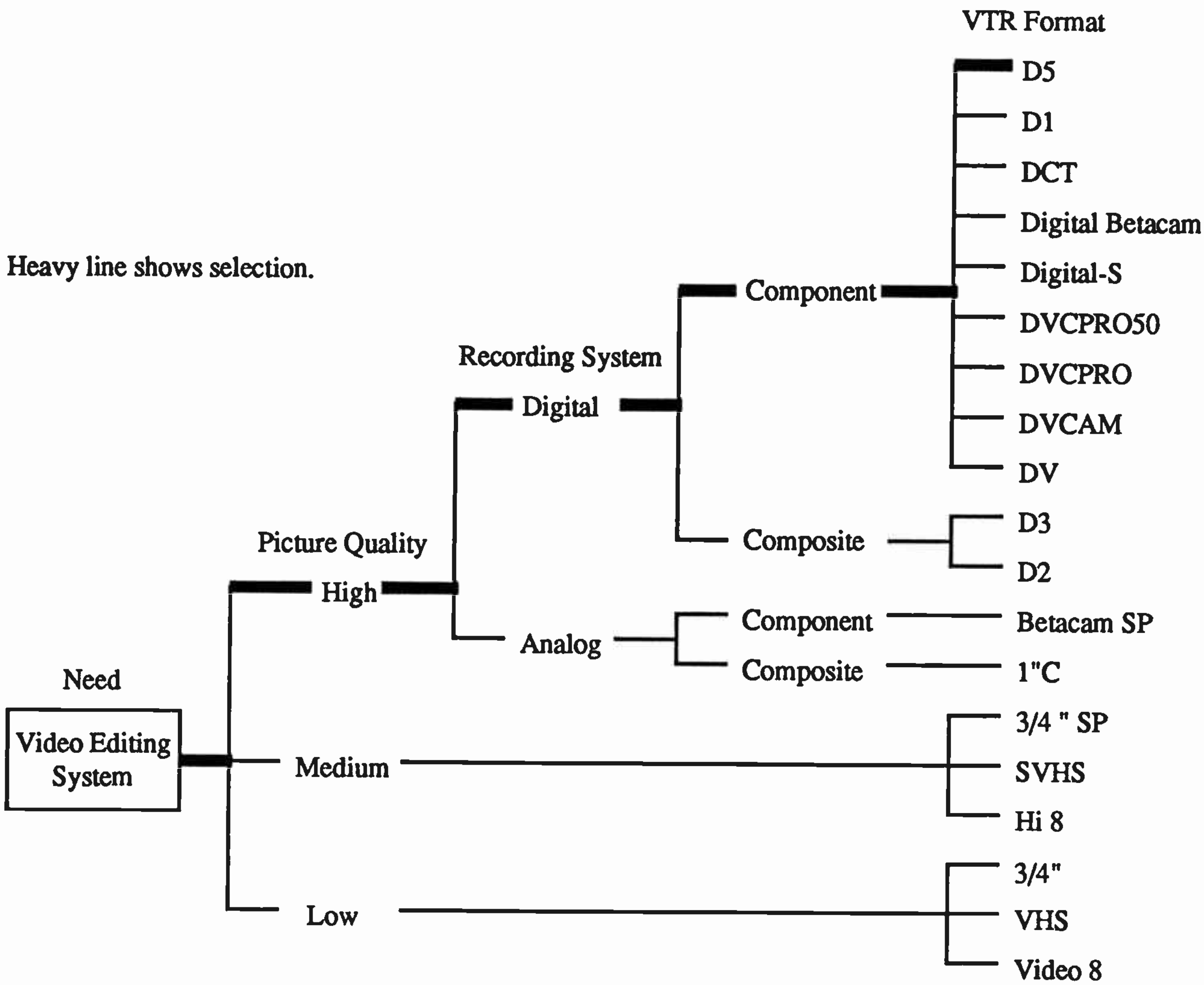


Figure 9.3-6. Trade tree example.

The criteria data on the candidates are then collected and tabulated on a decision analysis work sheet (see Figure 9.3-7). The attributes and limitations are listed in the first column, and the data for each candidate are given in adjacent columns to the right. The performance data are available from vendor spec sheets or may require laboratory testing and analysis to determine. Each attribute is given a relative score from 1 to 10 based on its comparative performance relative to the other candidates. Utility function graphs can be used to assign logical scores for each attribute (see Figure 9.3-8). The utility curve represents the advantage rating for a particular value of an attribute. A graph is made of ratings on the y axis vs. attribute value on the x axis, and specific scores that correspond to particular performance values can then be applied. The shape of the curve may take into account requirements, limitations and any other factor that will influence its value regarding the particular criteria being evaluated. The limits to which the curves should be extended run from the minimum value below which no further benefit will accrue to the maximum value above which no further benefit will accrue.

The scores are filled in on the decision analysis work sheet and multiplied by the weights to calculate the weighted score. The total of the weighted scores for each candidate then determines their ranking. As a rule, at least a 10% difference in score is acceptable as meaningful.

Further analysis can be applied in terms of evaluating the sensitivity of a decision to changes in the value of attributes, weights, subjective estimates and cost. Scores should be checked to see if changes in weights or scores would reverse the choice. How sensitive is the decision to changes in the system requirements or technical capabilities?

A trade table can be prepared to summarize the selection results (see Figure 9.3-9). Pertinent criteria

ALTERNATIVES:		Candidate 1			Candidate 2			Candidate 3		
WANTED	WT		SC	WT SC		SC	WT SC		SC	WT SC
Video Bandwidth (MHz)	10	5.6	10	100	6.0	10	100	5.0	9	90
Signal-to-Noise Ratio (dB)	10	60	8	80	54	6	60	62	10	100
10 bit quantizing	10	yes	1	10	yes	1	10	yes	1	10
Max program length (Hrs)	10	2	2	20	3	3	30	1.5	1.5	15
Read before write capable	5	yes	1	5	yes	1	5	no	0	0
Audio pitch correction avail	5	yes	1	5	no	0	0	yes	1	5
Capable of 16:9 aspect ratio	10	no	0	0	yes	1	10	yes	1	10
Employs compression	-5	yes	1	-5	no	0	0	yes	1	-5
SDI input/output	10	yes	1	10	yes	1	10	yes	1	10
Current installed base	8	medium	2	16	low	1	8	low	1	8
TOTAL WEIGHTED SCORE:				241			234			243

Figure 9.3-7. Decision analysis worksheet example. (Adapted from *System Engineering Management Guide*, Contract No. MDA 903-82-C-0339, Defense Systems Management Collage, Virginia, 1983.)

are listed for each alternative solution, which may be described quantitatively (high, medium or low).

Finally, the results of the trade study are documented in a report that discusses the reasons for the selections and which may also include the trade tree and the trade table.

Having a formal system of change control throughout the systems engineering process will keep all parties informed and prevent changes from being made without proper review and approval. Change control ensures that all documentation is kept up to date, and it can even help eliminate redundant documents. Finally, change control helps control project costs.

Description of System Elements

There are five categories of interacting system elements: equipment (hardware), software, facilities, personnel and procedural data. Performance, design and test requirements must be specified and documented for equipment, components and computer software elements of the system. Environmental and interface design requirements necessary for proper functioning of system elements within a facility must also be specified.

The documentation produced by the systems engineering process controls the evolutionary development of the system. Figure 9.3-10 illustrates the special purpose documentation used by one organization for each step of its systems engineering process.

The requirements are formalized in written specifications. By establishing clear standards for producing specifications, organizations can help reduce the variability of technical content and improve product quality. While the product should not be over-specified to the point of making the design too costly, requirements should also not be so vague or general that the product fails to meet customer needs. In large departmentalized organizations, a commitment to schedules can help ensure that other members of the organization can coordinate their time.

The systems engineering process does not actually design the system but rather produces the documentation necessary to define, design, build and test it. The technical integrity provided by this documentation ensures that the design requirements for the system elements reflect the functional performance requirements, that all functional performance requirements are satisfied by the combined system elements and that such requirements are optimized with respect to system performance requirements and constraints.

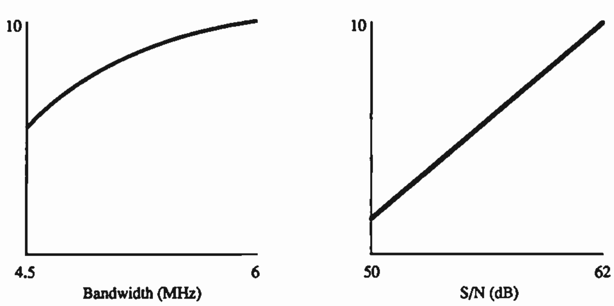


Figure 9.3-8. Attribute utility trade curve examples.

CRITERIA	Cool room only. Only normal convection cooling within enclosures	Forced cold air ventilation through rack then directly into return	Forced cold air ventilation through rack, exhausted into the room, then returned thru the normal plenum
Cost	Lowest. Conventional central air conditioning system used.	High. Dedicated ducting required. Separate system required to cool room	Moderate. Dedicated ducting required for input air.
Performance Equipment Temperature Room Temperature	Poor 80 to 120°F+ 65 to 70°F typical as set	Very Good 55 to 70°F typical 65 to 70°F typical as set	Very Good 55 to 70°F typical 65 to 70°F typical as set
Control Control of equipment temperature	Poor. Hot spots will occur within enclosures.	Very Good.	Very Good When the thermostat is set to provide a comfortable room temperature, the enclosure will be cool inside.
Control of room temperature	Good. Hot spots may still exist near power hungry equipment.	Good.	Good. If the enclosure exhaust air is comfortable for operators, the internal equipment must be cool.
Operator comfort	Good	Good. Separate room ventilation system required can be set for comfort.	Good. When the thermostat is set to provide a comfortable room temperature, the enclosure will be cool inside.

Figure 9.3-9. Trade table example.

Phases of a Typical System Design Project[1]

In the following section, the design of a complex modern video facility is used to illustrate the systems engineering approach.

Design Development

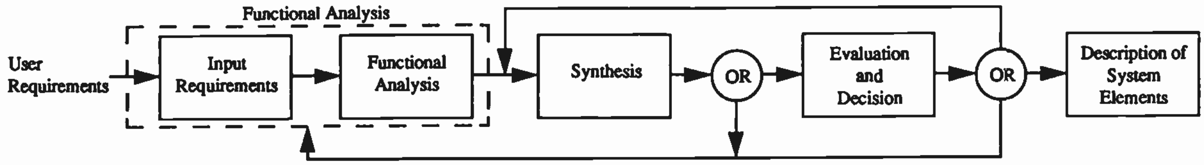
System design is carried out in a series of steps that leads to an operational facility. The bulk of the creative work, including research and preliminary design, are completed in the first phase of the project—the design development phase. During this phase, all project requirements are fully delineated and any constraints identified. Based on initial concepts and information, the design requirements are modified until all concerned parties are satisfied and approval is given for the final design work to proceed. The first objective of this phase is to answer the following questions:

- What are the functional requirements of the product of this work?
- What are the physical requirements of the product of this work?
- What are the performance requirements of the product of this work?
- Are there any constraints limiting design decisions?
- Will existing equipment be used?
- Is the existing equipment acceptable?
- Will this be a new facility or a renovation?
- Will this be a retrofit or upgrade to an existing system?
- Will this be a stand alone system?

In the design development phase, the equipment and functional requirements of each of the major technical areas of the facility are identified. In the case of facility renovation, the systems engineer's first order of business is to analyze existing equipment. A visit is made to the site to gather detailed information about the existing facility. Usually confronted with a mixture of acceptable and unacceptable equipment, the systems engineer must sort out which equipment meets current standards and which items should be replaced. Then, after soliciting input from the facility's technical and operational personnel, the systems engineer develops a list of needed equipment.

One of the systems engineer's most important contributions is the ability to identify and meet the needs of the "customer" and do it within the project budget. Based on the customer's initial concepts and any sub-

[1] Portions of this section were adapted from J.C. Whitaker, G. DeSantis and C. Paulson, *Interconnecting Electronic Systems*, CRC Press, Boca Raton, Fla., Chapter 1, 1992.



	Functional Flow Block Diagrams (FFBD)	Requirements Allocation Sheets (RAS)	Concept Description Sheets (CDS)	Trade Study Report (TSR)	Design Sheet (DS)
Basic Documents	Identify and sequence functions that must be accomplished to achieve system or project objectives. Develop the basics for establishing intersystem functional interfaces and identify system relationships.	Define the requirements and constraints for each of the functions and relate each requirement to the system elements of: a. Equipment b. Facilities C. Personnel d. Procedural data e. Computer software Time Line Sheets (TLS) Present critical functions against a time base in the required sequence of accomplishment.	Constrain the design to stop at a point in the cycle and create, at the gross level, a design or synthesis meeting the FFBD, RAS, TLS, requirements and constraints. Schematic Block Diagrams (SBD) Develop and portray schematic arrangement of system elements to satisfy system requirements.	Select, evaluate and optimize promising or attractive concepts. Document the tradeoff and supporting rationale. Consider all possible solutions within the framework of requirements.	Define, describe and specify performance, design and test criteria for the system elements a. Equipment b. Facilities c. Personnel d. Procedural data e. Computer software
Special Documents	End Item Maintenance Sheet (EIMS), Test Requmt Sheet (TRS), Production Sheets (PS), Logistic Support Analysis Record (LSAR) Identify maintenance, test, and production functions on a specific end item, subassembly, and component basis.				

Figure 9.3-10. Basic and special purpose documentation for systems engineering. (Adapted from Hoban, F. T., and Lawbaugh, W. M., *Readings in Systems Engineering*, NASA, Washington, DC, pg. 10, 1993.)

sequent equipment utilization research conducted by the systems engineer, the desired capabilities are identified as precisely as possible. Design parameters and objectives are also defined and reviewed. Functional efficiency is maximized to allow operation by a minimum number of personnel. Future needs are also investigated at this time, and future technical systems expansion is considered.

After the customer approves the equipment list, preliminary system plans are drawn up for review and further development. If architectural drawings of the facility are available, they can be used as a starting point for laying out an equipment floor plan. The systems engineer uses this floor plan to be certain adequate space is provided for present and future equipment as well as adequate clearance for maintenance and convenient operation. Equipment identification is then added to the architect's drawings.

Documentation should include, but not be limited to:

- Equipment list with prices
- Technical system functional block diagrams
- Custom item descriptions
- Rack and console elevations
- Equipment floor plans

The preliminary drawings and other supporting documents are prepared to record design decisions and to illustrate the design concepts for the customer. Renderings, scale models or full-size mock ups may also be needed to better illustrate, clarify or test design ideas.

Since ideas and concepts have to be exchanged and understood by all concerned parties, good communication skills are essential for team members. By completion of this phase of the project, the physical layout—the look and feel—and the functionality of the facility all will have been decided and agreed upon. If the design concepts appear feasible and the cost is within the anticipated budget, management can authorize work to proceed on the final detailed design.

ELECTRONIC SYSTEMS DESIGN

In a technical facility project, performance standards and specifications for such things as signal quality, stability, reliability and accuracy have to be established up-front. These will determine an acceptable performance level for equipment used in the system and affect the size of the budget. Access and processor speeds are important parameters when dealing with computer driven products. The systems engineer has to confirm whether or not selected equipment conforms to the standards.

At this point it must be determined what functions each system component will be required to fulfill, and how each will function with other components in the system. The management and operations staff usually know what they would like the system to do and how best to accomplish it. In fact, they have probably selected the equipment to do the job. With a familiarity of the capabilities of different equipment, the systems engineer should be able to contribute to this function-definition stage of the process. Questions that need to be answered include:

- What functions must be available to the operators?
- What functions are secondary and therefore not necessary?
- What level of automation should be required to perform a function?
- How accessible should the controls be?

Over-engineering or over-design by engineers and company staff should be avoided when planning technical system requirements. A staff member may, for example, ask for a seemingly simple feature or capability without fully understanding its complexity or the additional cost burden it may impose on a project. Other portions of the system may have to be compromised in order to implement the additional feature. An experienced systems engineer will be able to spot this and determine if the tradeoffs and added engineering time and cost are really justified.

When existing equipment is going to be used, an inventory list should be made as the starting point for developing a final equipment list. Confronted with a mixture of acceptable and unacceptable equipment, the systems engineer usually must sort out what meets current standards and what should be replaced. Then, after soliciting input from facility technical personnel, the systems engineer develops a summary of equipment needs, including future acquisitions. One of the systems engineer's most important contributions is the ability to identify and meet these needs within the facility budget.

A list of major equipment is prepared as part of the systems engineering process. The systems engineer selects the equipment based on experience with the products and on customer preferences and, when feasible, includes existing equipment that can be reused. A number of considerations are discussed with the facility customer to arrive at the best product selection. Some of the major ones include:

- Budget restrictions
- Space limitations
- Performance requirements
- Ease of operation
- Flexibility of use
- Functions and features
- Past performance history
- Manufacturer support

The goal here is to specify equipment that meets the functional requirements of the project efficiently and economically. Simplified block diagrams for the video, audio, control, data and communication systems are drawn, discussed with the customer and presented for approval.

Detailed Design

When the research and preliminary design development are completed, the details of the design are concluded. The design engineer prepares complete detailed documentation and specifications necessary for the fabrication and installation of the technical systems, including all major and minor components. Drawings must show the final configuration and relationship of each component to other elements of the system, as well as how they will interface with other building services, such as air conditioning and electrical power. This documentation must communicate the design requirements to the other design professionals, including the construction and installation contractors.

In this phase, the systems engineer develops final, detailed flow diagrams and schematics that show the interconnection of all equipment. Cable interconnection information for each type of signal is taken from the flow diagrams and recorded on the cable schedule. Cable paths are measured and timing calculations made for signals requiring synchronization (i.e., video, synchronizing pulses, subcarrier and digital audio). These timed cable lengths are entered onto the cable schedule.

The flow diagram is a schematic drawing used to show the interconnections between all equipment that will be installed. It contains much more detail than a block diagram; every wire and cable must be included on the drawings. A typical flow diagram for a video production facility is shown in Figure 9.3-11.

The starting point for preparing a flow diagram varies depending on the information available from the design development phase of the project and on the similarity of the project to previous projects. If a similar system has been designed in the past, the diagrams from that project can be modified to include the equipment and functionality required for the new system. New models of the equipment can be shown in place of their counterparts on the diagram, and only minor wiring changes need to be made to reflect the new equipment connections and changes in functional requirements. This method is efficient and easy to complete.

If the facility requirements do not fit any previously completed design, the block diagram and equipment list are used as a starting point. Essentially, the block diagram is expanded and details added to show all of the equipment and their interconnections as well as to show any details necessary to describe the installation and wiring completely.

An additional design feature that might be desirable for specific applications is the ability to easily disconnect a rack assembly from the system and relocate it. This would be the case if the system were to be prebuilt at a systems integration facility and later moved to

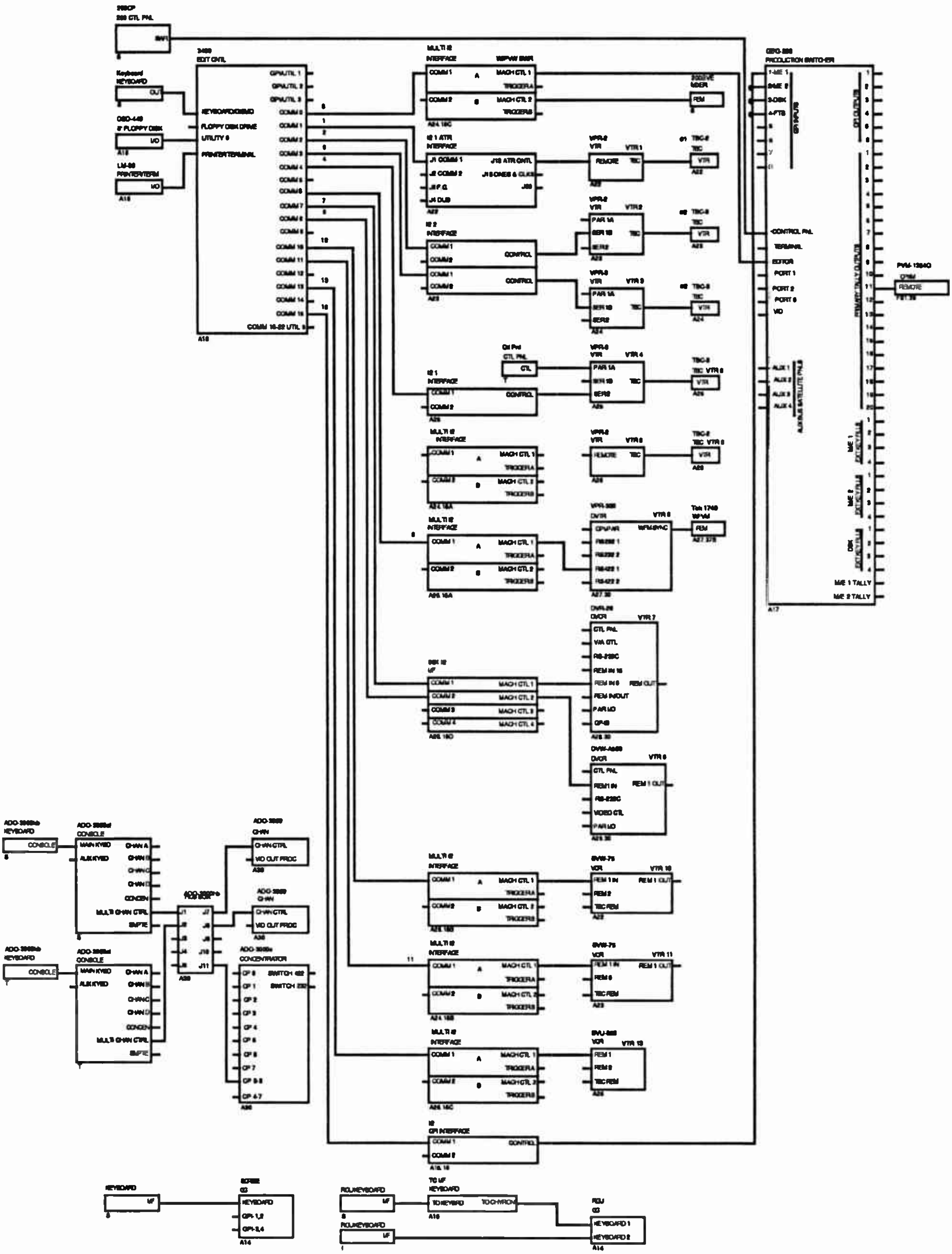


Figure 9.3-11. Video system control flow diagram.

and installed at the client's site. When this is a requirement, the interconnecting cable harnessing scheme must be well planned in advance and identified on the drawings and cable schedules.

Special custom items need to be defined and designed. Detailed schematics and assembly diagrams are drawn, parts lists and specifications are finalized and all necessary details worked out for these items. Mechanical fabrication drawings are prepared for consoles and other custom built cabinetry.

The design engineer provides the architect with layouts of cable runs and connections. Such detailed documentation simplifies equipment installation and facilitates future changes in the system. During preparation of final construction documents, the architect and the design engineer can firm-up the layout of the technical equipment wire ways, including access to flooring, conduits, trenches and overhead wire ways.

Dimensioned floor plans and elevation drawings are required to show placement of equipment, lighting, electrical cable ways, duct, conduit and HVAC ducting. Requirements for special construction, electrical, lighting, HVAC, finishes and acoustical treatments must be prepared and submitted to the architect for inclusion in the architectural drawings and specifications. This type of information, along with cooling and electrical power requirements, also must be provided to the mechanical and electrical engineering consultants (if used on the project) so they can begin their design calculations.

Equipment heat loads are calculated and submitted to the HVAC consultant. When locating equipment, care is taken to avoid excessive heat buildup within the equipment enclosures while still maintaining a comfortable environment for the operators.

Electrical power loads are calculated and submitted to the electrical consultant, and steps are taken to provide for sufficient power and proper phase balance.

Customer Support

The systems engineer can assist in purchasing equipment and help coordinate the move to a new or renovated facility. This can be critical if a great deal of existing equipment is being relocated. In the case of new equipment, the customer will find the systems engineer's knowledge of prices, features and delivery times to be invaluable. A good systems engineer will see to it that equipment arrives in ample time to allow for sufficient testing and installation. A good working relationship with equipment manufacturers helps guarantee their support and speedy response to the customer's needs.

The systems engineer can also provide engineering management support during planning, construction, installation and testing to help qualify and select contractors, resolve problems, explain design requirements and assure quality workmanship by contractors and technical staff.

The procedures described in this chapter outline an ideal scenario. In reality, management may often try to bypass many of the foregoing steps to save money, thinking this will eliminate unnecessary engineering costs and allow construction to start right away. Although a small company may attempt to handle the job by using in-house personal rather than outside professionals, this puts an added burden on staff who are already working full time taking care of the daily operation of the facility. With inadequate design detail and planning, which may result by using unqualified people, the job of setting technical standards and making the system work then defaults to the construction contractors, in-house technical staff or the installation contractor. This can result in costly and uncoordinated work-arounds and, of course, delays and added costs during construction, installation and testing. It makes the project less manageable and less likely to be completed successfully.

When large amounts of money and other resources are involved, management is well advised to recruit the services of qualified system engineers.

Budget Requirements Analysis

The need for a project may originate with customers, management, operations staff, technicians or engineers. In any case, some sort of logical reasoning or a specific production requirement will justify the cost. On small projects, like the addition of a single piece of equipment, it is easy to determine how much it will cost to purchase and install the equipment. On large projects, it is not always immediately apparent how much the project will cost. The project has to be analyzed by dividing it up into its constituent elements:

- Equipment
- Materials
- Resources (including money and staff needed to complete the project)

An executive summary or capital project budget request should contain a detailed breakdown of these elements to give management the information it needs to determine the return on investment and to make an informed decision on whether or not to authorize the project. A capital project budget request containing the minimum information might consist of the following items:

- Project name. (Use a name that describes the result of the project, such as "control room upgrade")
- Project number (if required). (A large organization that does many projects will use a project numbering system of some kind, or it may use a budget code assigned by the accounting department)
- Project description. (A brief description of what the project will accomplish, such as "design the technical system upgrade for the renovation of production control room 2")
- Initiation date. (The date the request will be submitted)
- Completion date. (The date the project will be completed)
- Justification. (The reason the project is needed)

- Material cost breakdown. (A list of equipment, parts and materials required for construction, fabrication and installation of the equipment)
- Total material cost
- Labor cost breakdown. (A list of personnel required to complete the project, their hourly pay rates, the number of hours they will spend on the project and the total cost for each)
- Total project cost. (The sum of material and labor costs)
- Payment schedule. (Estimation of individual amounts that will have to be paid out during the course of the project and the approximate dates each will be payable)
- Preparer's name and the date prepared
- Approval signature(s) and date(s) approved

More detailed analysis, such as return on investment, can be carried out by an engineer, but financial analysis should be left to accountants or other personnel with access to company financial data.

Feasibility Study and Technology Assessment

A feasibility study is necessary to evaluate new technology under consideration, including its ability to perform a desired function. The systems engineer may be called upon to assess the state-of-the-art in order to develop a new application. An executive summary or a more detailed report of evaluation test results may be required, in addition to a budget estimate, to help management make its decision.

Planning and Control of Scheduling and Resources

Several planning tools have been developed for planning and tracking project progress as well as scheduling and controlling resources. The most common ones are the Gantt Chart and the Critical Path Method (CPM) utilizing the Project Evaluation and Review (PERT) technique. Computerized versions of these tools have greatly enhanced management's ability to control large projects.

Project Tracking and Control

A standardized project status report form can provide consistent and complete information to the project manager regarding work completed, and money spent on resources and materials. A project status report containing the minimum information might contain the following items:

- Project number (if required)
- Date prepared
- Project name
- Project description
- Start date
- Completion date (the date this part of the project was completed)
- Total material cost
- Labor cost breakdown
- Preparer's name

Change Control

After part or all of a project design has been approved and money allocated to build it, any changes may increase or decrease the cost. Factors that affect the cost include:

- Components and materials
- Resources, such as labor and special tools or construction equipment
- Costs incurred because of manufacturing or construction delays

A *change order request* or *change order* is a method of reporting changes to management and soliciting its approval to proceed with the change. A change order includes a brief description of the change, the reason for the change, a summary of the effect it will have on costs and what effect it will have on the project schedule.

Management will exercise its authority and approve or disapprove each change based upon its understanding of the cost and benefits, and its own perceived need for modification of the original plan. Therefore, it is important that the systems engineer provide as much information and explanation as necessary to management.

A change order form containing the minimum information might include the following items:

- Project number
- Date prepared
- Project name
- Preparer's name
- Description of the change
- Reason for the change
- Equipment and materials to be added or deleted
- Material costs or savings
- Labor cost breakdown
- Labor costs or savings
- Total cost of this change (increase or decrease)
- Impact on the schedule.

Program Management

The Defense Systems Management College in Fort Belvoir, Virginia, defines systems engineering as:

Systems engineering is the management function which controls the total system development effort for the purpose of achieving an optimum balance of all system elements. It is a process which transforms an operational need into a description of system parameters and integrates those parameters to optimize the overall system effectiveness.

Systems engineering is both a technical process and a management process. Both processes must be applied throughout a program for it to be successful. The persons who plan and carry out a project constitute the project team; its makeup will vary depending on the size of the company and the complexity of the project. It is management's responsibility to provide the necessary human resources to complete the project.

Executive Management

The executive manager is the person who can authorize a project and who bears ultimate responsibility for its success. This person can allocate funds and delegate authority to others to accomplish the task. This person's job is to get things done through other people by assigning group responsibilities, coordinating activities between groups and resolving group conflicts. The executive manager establishes policy, provides broad guidelines, approves the project master plan, resolves conflicts and assures project compliance with commitments.

Executive management delegates the project management functions and assigns authority to qualified professionals, allocates a capital budget for the project, supports the project team and establishes and maintains a healthy relationship with project team members.

Management is responsible for providing clear information and goals—up front—based upon its needs and initial research. Before initiating a project, the company executive should be familiar with daily operation of the facility, analyze how the company works, how the staff do their jobs and what tools they need to accomplish the work. Some points that may need to be considered by an executive before initiating a project include:

- What is the current capital budget for equipment?
- Why does the staff currently use specific equipment?
- What function of the equipment is the weakest within the organization?
- What functions are needed but cannot be accomplished with current equipment?
- Are the staff satisfied with current hardware?
- Are there any reliability problems or functional weaknesses?
- What is the maintenance budget and is it expected to remain steady?
- How soon must the changes be implemented?
- What is expected from the project team?

Unless the manager has made a systematic effort to evaluate all the obvious points about the facility requirements, the not-so-obvious points may be overlooked. Overall requirements must be broken down into their component parts. The planning should be as basic as possible, and ideas that have too many branches should be simplified. Only after answering the appropriate questions will the executive manager be ready to bring in expert project management and engineering assistance. If the company executive does not make a concerted effort to investigate the needs and problems of a facility thoroughly before consulting experts, the expert advice will be shallow and incomplete, no matter how good the engineer.

Engineers work with the information they are given. They put together plans, recommendations, budgets, schedules, purchases, hardware and installation specifications based upon the information they receive from interviewing management and staff. If the management and staff have failed to go through the planning, reflection and refinement cycle before those interviews, the company will likely waste time and money.

PROJECT MANAGER

Project management can be applied to any undertaking that has a specific end objective. It is an outgrowth of the need to accomplish large complex projects in the shortest possible time, within the anticipated budget and with the required performance and reliability. Project management is based upon the realization that modern organizations may be so complex as to preclude effective management using traditional organizational structures and relationships.

The project manager must be a competent systems engineer, accountant and manager. As a systems engineer, project managers must understand analysis, simulation, modeling, reliability and testing techniques. The project manager must be aware of state-of-the-art technologies and their limitations. As an accountant, the project manager must be aware of the financial implications of their decisions and know how to control them. As a manager, planning and control of schedules is an important part of controlling the costs of a project and completing it on time. Also, as a manager, a project manager must possess the skills necessary to communicate clearly and convincingly with subordinates and superiors to make them aware of problems and their solutions.

The project manager has the authority to carry out a project and has been given the legitimate right to direct the efforts of the project team members. The project manager's power comes from the acceptance and respect accorded by superiors and subordinates. The project manager has the power to act and is committed to group goals.

The project manager is responsible for getting the project completed properly, on schedule and within budget, by utilizing whatever resources are necessary to accomplish the goal most efficiently. The manager provides project schedule, financial and technical requirement direction and evaluates and reports on project performance. This requires planning, organizing, staffing, directing and controlling all aspects of the project.

In this leadership role, the project manager is required to perform many tasks including:

- Assembling the project organization
- Developing the project plan
- Publishing the project plan
- Setting measurable and attainable project objectives and performance standards
- Setting attainable performance standards
- Determining the appropriate scheduling tools (PERT, CPM, and/or GANTT) for the project
- Using the scheduling tools to develop and coordinate the project plan, which includes budget, resources, and project schedule
- Developing the project schedule and project budget
- Managing the budget

- Recruiting personnel for the project
- Selecting subcontractors
- Assigning work, responsibility and authority so team members can make maximum use of their abilities
- Estimating, allocating, coordinating and controlling project resources
- Dealing with unrealistic specifications and resource needs
- Deciding on the right level of administrative and computer support
- Training project members on how to fulfill their duties and responsibilities
- Supervising project members, giving them day-to-day instructions, guidance and discipline as required to fulfill their duties and responsibilities
- Designing and implementing reporting and briefing information systems or documents that respond to project needs
- Controlling the project

Basic project management practices that can improve the chances for success include:

- Securing the necessary commitments from top management to make the project a success
- Setting up an action plan that will be easily adopted by management
- Using a work breakdown structure that is comprehensive and easy to use
- Establishing accounting practices that help, not hinder, successful project completion
- Preparing project team job descriptions properly upfront to eliminate conflict later on
- Selecting project team members appropriately the first time.

After the project is under way, the project manager should follow these steps:

- Manage the project, but make the oversight reasonable and predictable
- Get team members to accept and participate in the plans
- Motivate project team members for best performance
- Coordinate activities so they are carried out in relation to their importance with a minimum of conflict
- Monitor and minimize interdepartmental conflicts
- Get the most out of project meetings without wasting the team's productive time. Develop an agenda for each meeting and start on time. Conduct one piece of business at a time. Assign responsibilities where appropriate. Agree on follow-up and accountability dates. Indicate the next step for the group. Set the time and place for the next meeting. End on time.
- Spot problems and take corrective action before it is too late
- Discover the strengths and weaknesses in project team members and manage them to get desired results
- Help team members solve their own problems
- Exchange information with subordinates, associates, superiors and others about plans, progress and problems
- Make the best of available resources
- Measure project performance
- Determine, through formal and informal reports, the degree to which progress is being made
- Determine causes of and possible ways to act upon significant deviations from planned performance
- Take action to correct an unfavorable trend or to take advantage of an unusually favorable trend
- Look for areas where improvements can be made
- Develop more effective and economical methods of managing
- Remain flexible
- Avoid *activity traps*
- Practice effective time management

The project manager should make sure their subordinates:

- Comprehend what they are supposed to do, preferably in terms of an end product
- Have a clear understanding of what their authority is and its limits
- Understand their relationships with others
- Recognize what constitutes a job well done in terms of specific results
- Appreciate when and what they are doing exceptionally well
- Are shown concrete evidence that there are just rewards for work well done and for work exceptionally well done.
- Comprehend where and when they are falling short of expectations
- Are made aware of what can and should be done to correct unsatisfactory results
- Feel that their superior has an interest in them as individuals, believes in them and is anxious for them to succeed and progress.

By fostering good relationships with associates, the project manager will have less difficulty communicating with them. The fastest, most effective communication takes place among people with common points of view.

By watching what is going on. the competent project manager can perceive problems long before they flow through the paper system. A project manager who spends most of their time in the management office instead of roaming through the places where the work is being done is headed for disaster.

Systems Engineer

Systems engineer means different things to different people. The systems engineer is distinguished from the engineering specialist, who is concerned with only one aspect of a well defined engineering discipline. By contrast, the systems engineer must be able to adapt to the requirements of almost any type of system. The systems engineer provides a wealth of experience gained from successful approaches to technical problems and developed through hands-on exposure to a variety of situations. A professional with knowledge, experience and skills in a specialized and learned field or fields, the systems engineer is highly trained in analyzing problems and developing solutions that satisfy management objectives. They take data from the overall development process and, in return, provide data in the form of requirements and analysis results to the process.

Education in electronics theory is a prerequisite for designing systems that employ electronic components. As a graduate engineer, the systems engineer has the education required to design electronic systems correctly. Mathematics skill acquired in engineering school is one of the tools used by the systems engineer to formulate solutions to design problems and analyze test results. Knowledge of testing techniques and theory enables the specification of system components and performance and measurement of results. Drafting and writing skills are required for efficient preparation of the necessary documentation needed to communicate the design to technicians and contractors who will have to build and install the system.

A competent systems engineer has a wealth of technical information available for speeding up the design process and helping make cost effective decisions. If necessary information is not at hand, the systems engineer knows where to find it. The experienced systems engineer is familiar with proper fabrication, construction, installation and wiring techniques and can spot and correct improper work.

Training in personnel relations, a part of the engineering curriculum, helps the systems engineer communicate and negotiate professionally with subordinates and management.

Small in-house projects can be completed on an informal basis and, indeed, this is probably the normal routine when projects are simple and uncomplicated. On large projects, however, the systems engineer's involvement usually begins with preliminary planning and continues through fabrication, implementation and testing. The degree to which program objectives are achieved is an important measure of the systems engineer's contribution.

During the design process, the systems engineer:

- Concentrates on results and focuses work according to management objectives
- Receives input from management and staff
- Researches the project and develops a workable design
- Assures balanced influence of all required design specialties
- Conducts design reviews
- Performs trade-off analyses
- Assists in verifying system performance
- Resolves technical problems related to the design, interface between system components and integration of the system into any facility

Aside from designing a system, the systems engineer has to answer questions and resolve problems that may arise during hardware fabrication and installation and must also monitor the quality and workmanship of the installation. The systems engineer also has to be concerned with the testing and calibration of hardware and software upon completion.

During the production or fabrication phase, the systems engineer is concerned with:

- Verifying system capability and system performance
- Maintaining the system baseline
- Forming an analytical framework for producibility analysis

Depending on the complexity of the new installation, the systems engineer may have to provide orientation and operating instruction to the users. During the operational support phase, systems engineers:

- Receive input from users
- Evaluate proposed changes to the system
- Establish their effectiveness
- Facilitate the effective incorporation of changes, modifications and updates.

Depending on the size of the project and the management organization, the systems engineer's duties will vary. In some cases the systems engineer may have to assume the responsibilities of planning and managing smaller projects.

Other Project Team Members

Other key members of the project team where building construction may be involved include the following:

- Architect—responsible for design of the structure
- Electrical engineer—responsible for power system design if not handled by the systems engineer
- Mechanical engineer—responsible for HVAC and plumbing designs
- Structural engineer—responsible for concrete and steel structures
- Construction contractors—responsible for executing the plans developed by the architect and mechanical and structural engineers

Other outside contractors may be employed for certain specialized custom items that can not be developed or fabricated internally or by any of the other contractors.

CONCLUSION

Systems theory is the theoretical basis of systems engineering, which is an organized approach to the design of complex systems. The key components of systems theory applied in systems engineering are a holistic approach, the decomposition of problems, the exploitation of analogies and the use of models.

A formalized technical project management technique used to define systems includes three major steps:

- Defining the requirements in terms of functions to be performed as well as the measurable and testable requirements that describe how well each function must be performed

- Synthesizing a way to fulfil the requirements
- Studying the tradeoffs and selecting one solution from the alternative solutions.

In the final analysis, beyond any systematic approach to systems engineering, engineers have to engineer. Former NASA Administrator Robert A. Frosch, in a speech to a group of engineers in New York, urged a common sense approach to systems engineering:

Systems, even very large systems, are not developed by the tools of systems engineering, but only by the engineers using the tools. . . . I can best describe the spirit of what I have in mind by thinking of a music student who writes a concerto by consulting a check list of the characteristics of the concerto form, being careful to see that all of the canons of the form are observed, but having no flair for the subject, as opposed to someone who just knows roughly what a concerto is like, but has a real feeling for music. The results become obvious upon hearing them. The prescription of technique cannot be a substitute for talent and capability . . .

BIBLIOGRAPHY

Defense Systems Management, *Systems Engineering Management Guide*, Defense Systems Management College, Fort Belvoir, VA, 1983.

Delatore, J.P., E.M Prell, and M.K. Vora, "Translating Customer Needs Into Product Specifications," *Quality Progress*, January 1989.

Finkelstein, L., "Systems Theory," *IEEE Proceedings*, vol. 135, part A, no. 6, pp. 401–403, July 1988.

Hoban, F.T., and W.M. Lawbaugh, *Readings In Systems Engineering*, NASA Science and Technical Information Program, Washington, DC, 1993.

Shinners, S. M., A *Guide to Systems Engineering and Management*, Lexington Books, Lexington, MA, 1976.

Tuxal, J. G., *Introductory System Engineering*, McGraw-Hill, New York, NY, 1972.

9.4
HUMAN EXPOSURE TO RADIO FREQUENCY FIELDS

KELLY T. WILLIAMS
NATIONAL ASSOCIATION OF BROADCASTERS, WASHINGTON, DC

Since the last edition of this handbook was published the Federal Communications Commission has made comprehensive changes to its guidelines for human exposure to radio frequency fields. These changes were adopted in the *Report and Order* in ET Docket 93-62, released August 1, 1996. Guidance on complying with these new regulations is provided by the FCC in *OET Bulletin 65, Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Fields*, and in *Supplement A* to this bulletin, *Additional Information for Radio and Television Broadcast Stations*.

The most significant change incorporated into the new rules is the institution of a two-tiered table of human exposure limits. The first tier applies to environments where the general population is exposed to radio frequency fields and the second tier applies to environments where only workers who are fully aware of, and able to control, their potential for exposure are exposed.

On October 15, 1997, all applications to the FCC for construction permits, license renewals and station modifications became subject to analysis under the new RF exposure standards. All facilities constructed after October 15, 1997, were required to comply with the new standards even if the construction permit authorizing their construction was granted and/or applied for before this deadline.

Existing stations that were in compliance with the old RF exposure standards on October 15, 1997, are not required to bring their facilities into compliance with the new standards until September 1, 2000, unless they file a license renewal or station modification application prior to that date. If they are not in compliance by September 1, 2000, they must file an Environmental Assessment with the FCC on that date in accordance with the requirements in Part 1, Subpart I of the FCC Rules.

Compliance with the FCC's RF exposure guidelines is the responsibility of FCC licensees. Even though tower owners are now primarily responsible for compliance with the FCC's tower painting and lighting requirements, FCC licensees remain responsible for ensuring that the emissions from their antennas do not result in excessive levels of human exposure to RF energy. If a tower owner will have maintenance workers climbing on a tower, tenant licensees must cooperate with the tower owner in controlling the workers' exposure to RF energy.

Because the FCC RF exposure guidelines are *human exposure* guidelines, they apply only to locations that are accessible to humans. If a particular antenna produces an RF field that is above the FCC limits, but it only produces this field in a location that is never occupied by humans, then the antenna does not pose an RF exposure hazard according to the FCC guidelines. Also, the FCC RF exposure guidelines are based on the average level of exposure over a fixed period of time (six minutes for the occupational/controlled exposure limits, 30 minutes for the general population/uncontrolled limits). It is therefore possible for a person to be exposed to a level of RF energy that exceeds the FCC average limit for a brief period of time, say one minute, yet not be exposed to a level of RF energy that exceeds the FCC guidelines because their average exposure over the 6- or 30-minute period falls below the FCC limit. This might be the case, for example, if a transmitter were located near a roadway and the vehicles passing through an area of excessive RF energy were only in the area for a second or two. Of course, in such a case, a person could be exposed to a level of RF energy that exceeds the FCC limit if their vehicle were to come to a stop within the area of excessive RF energy and remain there long enough for their average exposure to exceed the FCC guidelines. As illustrated in this example, it is often not possible to control exposure to the extent that averaging times can be applied, especially with regard to the general population. When it comes to general population exposure, it is usually necessary to assume continuous exposure. The exposure guidelines specified by the FCC are shown in Table 9.4-1.

According to FCC rules, if an area around an antenna produces RF fields that exceed the exposure guidelines in Table 9.4-1, the actions that must be taken by the licensee will depend on whether or not the area of high exposure will be occupied only by people who are fully aware of the potential for exposure and can exercise control over their exposure. If the area will be occupied by anyone who is not fully aware of the potential for exposure, or by anyone who cannot exercise control over their exposure, then the more restrictive general population/uncontrolled exposure limits apply.

If the area will only be occupied by workers who understand RF exposure hazards then the licensee must ensure that the appropriate steps are taken to prevent workers in the vicinity of the high exposure area from

Table 9.4-1
FCC RF Exposure Guidelines

(a) FCC RF Exposure Standard for Occupational/Controlled Exposure

Frequency Range (MHz)	Electric Field Strength (E) (V/m)	Magnetic Field Strength (H) (A/m)	Power Density (S) (mW/cm^2)	Averaging Time $\|E\|^2$, $\|H\|^2$ or S (minutes)
0.3–3.0	614	1.63	100	6
3.0–30	1842/(freq. in MHz)	4.89/(freq. in MHz)	900/(freq. in MHz)^2	6
30–300	61.4	0.163	1.0	6
300–1,500	—	—	(freq. in MHz)/300	6
1,500–100,000	—	—	5	6

Notes:

V/m = volts per meter
A/m = amperes per meter
mW/cm^2 = milliwatts per square centimeter

The occupational/controlled limits apply in situations where people are exposed to RF energy as a consequence of their employment, provided these people are fully aware of the potential for exposure and can exercise control over their exposure. These limits also apply in situations where any person is passing through the RF field and is made aware of the potential for exposure.

(b) FCC RF Exposure Standard for General Population/Uncontrolled Exposure

Frequency Range (MHz)	Electric Field Strength (E) (V/m)	Magnetic Field Strength (H) (A/m)	Power Density (S) (mW/cm^2)	Averaging Time $\|E\|^2$, $\|H\|^2$ or S (minutes)
0.3–1.34	614	1.63	100	30
1.34–30	824/(freq. in MHz)	2.19/(freq. in MHz)	180/(freq. in MHz)^2	30
30–300	27.5	0.073	0.2	30
300–1,500	—	—	(freq. in MHz)/1,500	30
1,500–100,000	—	—	1.0	30

Notes:

V/m = volts per meter
A/m = amperes per meter
mW/cm^2 = milliwatts per square centimeter

The general population/uncontrolled exposure limits apply in situations where the general public may be exposed. They also apply in situations where people are exposed as a consequence of their employment and may either not be fully aware of the potential for exposure or unable to exercise control over their exposure.

being exposed to RF levels in excess of the occupational/controlled exposure limits. Generally, this means identifying and marking the area where the levels of RF energy exceed the occupational/controlled limits, and taking the necessary steps to ensure that when workers must enter the area, they are not exposed to RF levels in excess of the occupational/controlled limits. This can be accomplished by reducing the transmitter's output power or turning the transmitter off. It can also be accomplished by limiting the amount of time that workers may spend in the hazardous area and/or requiring the exposed workers to wear RF protective clothing that provides effective RF shielding over the appropriate frequency range.

If an area around or near a tower contains RF fields that exceed the general population/uncontrolled exposure limits in Table 9.4-1, *and* the area is accessible to either a) workers who may not be fully aware of the potential for exposure or b) the general public, then measures must be taken to make the area inaccessible to these people. Generally, these measures will involve the use of fences and warning signs. Fences and/or warning signs are required, even if the area that contains the dangerous RF fields is located entirely on private property (licensees must also protect *trespassers* from exposure to dangerous RF fields).

RF HOT SPOT

An RF hot spot is a localized area where the RF field strength is substantially higher than the surrounding area. An RF hot spot is generally caused by one of three factors: 1) the intersection of multiple beams of RF energy; 2) standing waves produced by reflective surfaces, such as the ground; and 3) induced currents in conductive objects exposed to ambient RF fields. Broadcasters are responsible for reducing the radiation at RF hot spots to within the FCC RF exposure limits.

Hot spots that could expose the general public to levels of RF radiation higher than the FCC limits are usually the result of induced currents in conductive objects. Conductive objects that may cause RF hot spots include, but are not limited to, metal playground equipment, storage sheds, and heating/air conditioning ducts.

When confronted with a potential hot spot problem, a broadcaster should first determine if the level of RF exposure at the hot spot exceeds the FCC limits. It is generally advisable to hire a professional engineer to determine whether or not the suspected hot spot exceeds these limits.

If the measured exposure level *does not* exceed the FCC limits, the broadcaster is not obligated to take any corrective action. However, in the interest of good public relations, many broadcasters may want to alleviate the RF radiation problem anyway. If the measured exposure level *does* exceed the FCC limits, then the broadcaster must work with the property owner to develop a means for reducing the RF radiation. This could mean, for example, developing a grounding system for metal playground equipment, or adding nonconductive spacers to interior duct work.

Broadcasters are responsible for reducing the electromagnetic energy at RF hot spots to within the FCC limits *even if* the hot spot is the result of new building construction that was performed after the broadcaster commenced operations. In cases where the hot spot is the result of multiple RF signals from multiple broadcast (or other radio service) antennas, it is the shared responsibility of all transmitter licensees to correct the hot spot problem. However, in multiple transmitter situations, if any particular transmitter's contribution to the hot spot is less than 5% of the FCC limit, then the licensee of that transmitter is not required to participate in the correction of the hot spot problem.

Hot spots are often found in the immediate vicinity around radio towers because towers are in close proximity to the transmitting antennas. The metal tower structure is located in the strongest portion of the electromagnetic field produced by the tower's antenna(s), so the level of currents induced in the tower are generally at a maximum for the antenna(s) in question. These hot spots are of particular significance because radio towers often come in direct contact with human tissue when workers are climbing them, thus exposing the workers to the maximum level of energy that is produced by the hot spot. In these instances, broadcasters should take whatever steps are necessary to ensure worker safety.

Addressing RF hot spot problems is an important issue for broadcasters because Section 1.1307(b) of the FCC rules requires that, when filing applications for new stations, major modifications, and license renewals, broadcasters include information about the potential of their facilities to expose workers or the general public to levels of RF energy in excess of the FCC RF exposure guidelines—if such potential exists. A certification of RF exposure compliance can amount to a misrepresentation if, later in the license term, it is determined that a hot spot in the station's radiation pattern was resulting in excessive exposure to workers and/or the public, and the station was aware of this problem when it filed its application with the FCC.

FENCING AND SIGN POSTING REQUIREMENTS

The FCC requires fences around many AM towers to protect people near the towers from electrocution. Although the FCC does not specifically require fences around other towers, it does require licensees to restrict access to areas where there are excessive levels of electromagnetic energy. Constructing a fence is a very practical means of restricting access to these areas.

Certain types of AM towers—series fed, folded unipole, and insulated base—can shock or electrocute anyone who touches the tower while standing on the ground. These towers must be enclosed within a locked fence to prevent human contact. Appropriate warning signs must be posted on the outside of the fence facing all normal angles of approach. The fence must be designed to allow authorized personnel ready access to each tower base at all times for meter reading and maintenance purposes.[1]

As noted above, FM and TV towers, and AM towers that do not have series fed, folded unipole, or insulated base antennas, are not specifically required by FCC rules to be fenced. However, the FCC does require broadcast stations to indicate, on their applications, that the radio frequency radiation of the facility will not have a significant environmental impact (*i.e.*, it will not expose *anyone* to radio frequency fields that exceed the FCC RF exposure limits described in Table 9.4-1). If a station cannot make such a statement, it must submit an Environmental Assessment to the FCC, and may be required to modify its facility or risk denial of its application (whether the application is for new construction, a modification to an existing facility, a license renewal, etc).

In order for a station to be able to indicate that its application will not have a significant environmental impact, it must 1) control workers' access to areas with RF fields that exceed the occupational/controlled exposure limits described in Table 9.4-1; and 2) restrict public access to areas with RF fields that exceed the general population/uncontrolled exposure limits described in Table 9.4-1. To restrict worker access to areas that exceed the occupational/controlled exposure limits, the posting of signs is generally sufficient. To restrict general public access to areas that exceed the general population/uncontrolled exposure limits, a fence must usually be constructed around the area *and* appropriate warning signs must be posted. If the station's antenna is located in an extremely remote location not likely to be visited by the public then only warning signs are required to restrict access—though fences are still recommended.

The FCC expects licensees to restrict general public access to RF fields that exceed the general population/uncontrolled exposure limits. It includes trespassers among those whose access must be restricted, since trespassers may be inclined to climb a tower to an area close enough to an antenna where the exposure limits are exceeded, therefore, NAB generally recommends that all broadcast towers be fenced.

Height of Fence

The FCC recommends that an antenna fence be high enough to prevent people from jumping over it. It

[1] See Section 73.49 of the FCC rules.

recognizes that people who want to climb a fence will usually do so, regardless of its height, and the fencing requirement is not intended to keep these individuals out.

Distance from Tower to Fence (AM Towers)

The fence should be located far enough from the tower so that anyone standing outside of the fence will not be exposed to radio frequency fields in excess of the FCC limits for general population/uncontrolled exposure. Because the areas where exposure to RF energy in excess of the FCC limits will be in the near-field of an AM broadcast antenna, compliance with *both* the electric *and* magnetic field limits must be evaluated. (In contrast, for FM and TV broadcast facilities, compliance at the base of the tower can generally be based on measurements of *either* the electric or magnetic field because compliance with one of these in the far field of the antenna will generally mean compliance with the other.) The Environmental Protection Agency (EPA) developed a computer program that was used by the FCC to predict how far away a person must be from an AM tower to ensure compliance with the FCC's limits for electric and magnetic field human exposure. Using the EPA program, the FCC produced four tables (one each for four different tower heights) that can be used to determine the required distance between the tower and the fence.[2] The FCC tables are reproduced here as Table 9.4-2 (a)–(d). For transmitter power values and/or tower heights that are not listed in Table 9.4-2 linear interpolation may be used.

The tower heights listed in Table 9.4-2 are the electrical heights. The electrical height of most AM towers is generally a little higher than the physical height because the velocity of propagation in a metal tower is less than in the air and is a function of the tower's cross-section.

Distance from Tower to Fence (FM Towers)

The fence around an FM tower should be located far enough from the tower so that anyone standing outside the fence will not be exposed to radio frequency fields in excess of the FCC limits for general population/uncontrolled exposure. If the FM antenna is high enough off the ground, compliance with these limits will be achieved anywhere on the ground, and a fence is not necessary. Table 9.4-3 shows the minimum height of the center of radiation for various *single* FM antennas that is necessary to keep locations on the ground from exceeding the FCC general population/uncontrolled exposure limits.[3] Readers should take particular note that the ERP values in Table 9.4-3 are for *total* effective radiated power. For example, a 6 kW

Table 9.4-2

Minimum Required Distances from Various AM Towers to Their Fences.

(a) 1/10-Wave Tower

	Transmitter Power (kW)			
Frequency (kHz)	50	10	5	1
535–740	13 m	7 m	6 m	3 m
750–940	12 m	7 m	5 m	3 m
750–1140	11 m	6 m	5 m	3 m
1150–1340	10 m	6 m	5 m	3 m
1350–1540	10 m	6 m	5 m	3 m
1550–1705	10 m	6 m	5 m	3 m

1 m = 3.28 feet

(b) 1/4-Wave Tower

	Transmitter Power (kW)			
Frequency (kHz)	50	10	5	1
535–740	4 m	2 m	2 m	1 m
750–940	4 m	2 m	2 m	1 m
950–1140	4 m	2 m	2 m	1 m
1150–1340	4 m	2 m	2 m	1 m
1350–1540	4 m	2 m	2 m	1 m
1550–1705	5 m	2 m	2 m	1 m

1 m = 3.28 feet

(c) 1/2-Wave Tower

	Transmitter Power (kW)			
Frequency (kHz)	50	10	5	1
535–740	4 m	3 m	2 m	2 m
750–940	4 m	2 m	2 m	2 m
950–1140	4 m	2 m	2 m	1 m
1150–1340	4 m	2 m	2 m	2 m
1350–1540	4 m	2 m	2 m	2 m
1550–1705	4 m	3 m	2 m	1 m

1 m = 3.28 feet

(d) 5/8-Wave Tower

	Transmitter Power (kW)			
Frequency (kHz)	50	10	5	1
535–740	4 m	3 m	2 m	1 m
750–940	4 m	2 m	2 m	1 m
950–1140	4 m	2 m	2 m	1 m
1150–1340	4 m	2 m	2 m	1 m
1350–1540	4 m	2 m	2 m	1 m
1550–1705	4 m	2 m	2 m	2 m

1 m = 3.28 feet

Class A station that is using a circularly polarized antenna with 6 kW horizontal ERP and 6 kW vertical ERP has a *total* ERP of 12 kW.

In order for a location *on the ground* to be compliant with the FCC standard, it must not exceed the FCC exposure limit anywhere from ground level up to 2 meters above ground level. This is because 2 meters is the approximate upper range of the height of a standing person. The antenna height above ground level figures in Table 9.4-3 take this 2 meter "standing person" factor into account.

If, according to Table 9.4-3, a particular FM antenna is close enough to the ground to create a *danger zone*—

[2] See FCC OET Bulletin No. 65, *Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Fields, Supplement A*, August 1997, pp. 1–5.

[3] See FCC OET Bulletin No. 65, *Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Fields, Supplement A*, August 1997, pp. 11–18.

Table 9.4-3
Minimum Height for Single FM Antenna Compliance with General Population/Uncontrolled Exposure Limits

TOTAL ERP (H+V) (kW)	2 bays	4 bays	6 bays	8 bays	10 bays	12 bays
0.5	9.1	9.0	11.5*	14.9*	18.3*	21.7*
	5.3	8.1*	11.5*	14.9*	18.3*	21.7*
3	19.3	19.2	19.0	18.8	18.7	21.7*
	10.1	8.1	11.5*	14.9*	18.3*	21.7*
10	33.6	33.3	33.1	32.7	32.4	32.1
	16.8	13.1	11.3	14.9*	18.3*	21.7*
25	51.9	51.5	51.1	50.6	50.0	49.6
	25.4	19.6	16.6	15.1	18.3*	21.7*
50	72.6	71.9	71.4	70.7	69.9	69.3
	35.1	26.8	22.7	20.6	19.1	17.6
75	88.4	87.7	87.0	86.1	85.1	84.4
	42.6	32.5	27.3	24.8	22.9	21.2
100	101.8	100.9	100.1	99.1	98.0	97.1
	48.8	37.1	31.2	28.3	26.2	24.1
125	113.6	112.6	111.7	110.5	109.3	108.4
	54.6	41.2	34.6	31.4	29.0	26.7
150	124.2	123.1	122.2	120.9	119.6	118.5
	59.6	45.0	37.9	34.2	31.6	29.1
175	133.4	132.8	131.8	130.4	129.0	127.9
	64.0	48.3	40.6	36.7	33.9	31.2
200	143.1	141.8	140.8	139.3	137.7	136.5
	68.0	51.7	43.3	39.1	36.1	33.2

(antenna heights are in meters)

Notes:

1. For each ERP/number-of-bays entry, the number on top represents a dipole antenna, which is the worst case. The number on the bottom represents a best-case, typically available modern antenna (i.e., one that has minimal radiation in the downward direction).
2. The data in this table applies to single FM antennas with supporting tower bases that are no lower than the surrounding terrain.
3. Interpolation can be used to obtain intermediate values.
4. The entries marked with an asterisk ("*") are cases where, if the antenna's center of radiation were located at the minimum safe height predicted using the FCC exposure standard, the lowest antenna element would actually be below ground level. Since this would clearly be unacceptable, these numbers have been adjusted upward to ensure that the bottom antenna element will be at least 3 meters above the ground, which is the approximate maximum height for a standing person plus one meter for safety.

where the levels of RF energy exceed the FCC limits for general population/uncontrolled exposure—the most appropriate thing to do is have a professional engineer measure the RF fields around the tower and produce a map illustrating the *danger zone*. This area must then be fenced and posted with RF radiation warning signs.

Distance from Tower to Fence (NTSC Analog TV Towers)

The fence around an NTSC television tower should be located far enough from the tower so that anyone standing outside the fence will not be exposed to radio frequency fields in excess of the FCC limits for general population/uncontrolled exposure. If the TV antenna is high enough off the ground, compliance with these limits will be achieved anywhere on the ground, and a fence is not necessary.

The FCC uses a formula developed by the EPA to predict the strength of RF fields at the base of television towers. This formula assumes that some of the transmitted energy is reflected off of the surface of the earth, thus increasing the level of RF energy at ground level over what would be expected without surface reflection. The formula is as follows:

Equation 1: Minimum Height for Single TV Antenna Compliance with FCC Exposure Limits

$$[MAH] = 2 + \sqrt{\frac{419.84F^2(0.4(VERP) + AERP)}{4\pi S}}$$

where:

MAH = minimum antenna height in meters (2 is added to ensure compliance at head level for a standing person)

F = typical relative field factor in the downward direction (*i.e.*, the amount of energy that the antenna radiates downward expressed as a fraction of the amount of radiation in its main beam. "Downward" is defined as any downward direction that is at an angle of at least 60°, and no more than 90°, with respect to the antenna's horizontal plane.

VERP = total (horizontal plus vertical) peak visual ERP in watts

AERP = total (horizontal plus vertical) aural ERP in watts

S = FCC power density limit for human exposure in $\mu W/cm^2$

Equation 1 can be used for both VHF and UHF stations. For VHF stations with *batwing* elements, the typical value for F is 0.20. The FCC general population/uncontrolled exposure limit at VHF frequencies is 200 $\mu W/cm^2$. For UHF stations, the typical value for F is 0.10, and the FCC general population/uncontrolled exposure limit is $0.67f$ $\mu W/cm^2$, where "f" is the center frequency of the UHF channel in MHz. If the manufacturer's data for a particular antenna includes a value for F, then that value should be used.

If, according to Equation 1, a particular TV antenna is close enough to the ground to create a *danger zone*—where the levels of RF energy exceed the FCC guidelines for general population/uncontrolled exposure—the most appropriate thing to do is have a professional engineer measure the RF fields around the tower and produce a map illustrating the *danger zone*. This area must then be fenced and posted with RF radiation warning signs.

Distance from Tower to Fence (ATSC Digital TV Towers)

When the FCC adopted its latest edition of *OET Bulletin 65*, it indicated that its model for predicting RF exposure hazards around television broadcast towers would have to be modified for use with digital television signals.[4] It is possible that this will occur in the future. However, in the meantime, it is appropriate to use Equation 1 for DTV towers with an appropriate relative field factor for the type of antenna being used.

WARNING SIGNS

Areas that are fenced to prevent people from being exposed to RF energy levels in excess of the FCC exposure limits must also be posted with RF hazard warning signs so that people will understand why access to the areas is restricted. The number and location of signs should be sufficient enough to be noticed from any normal angle of approach.

AM towers that are fenced to prevent shock or electrocution (series fed, folded unipole and insulated base), must have signs that warn of the high voltage danger. The number and location of signs should be sufficient enough to be noticed from any normal angle of approach.

If a series fed, folded unipole or insulated base AM antenna creates an area of excessive RF energy, signs must be posted to warn of *both* the excessive RF energy and the potential for electric shock. Broadcasters can purchase RF and high voltage warning signs from the NAB.

To purchase a sign contact:

NAB Services
1771 N Street, N.W.
Washington, DC 20036-2891
Tel: (800) 368-5644

MEASUREMENT PROCEDURES

Whenever practical, field measurement of the RF power density levels is to be preferred over calculation, as the definitive method of determining those levels. Care must be taken, however, when taking measurements for the purposes of determining compliance with the ANSI RF exposure guidelines, in order that the examination of the site be complete. The principal concern when taking these measurements is to identify all areas requiring access restrictions, in order to protect fully the liability exposure of the FCC licensee(s) at the site. Failure to identify these areas begs the question of licensee candor in self-certification to the FCC at license renewal, as well as the issue of negligence in the event of legal action based upon "excessive" exposure to RF energy. It is important that the measurement process should be accurately described to allow for a high degree of repeatability in future measurements by others.

Physical Parameters

Power density is most commonly expressed in units of milliwatts per square centimeter (mW/cm^2). Instruments designed for electromagnetic field hazard surveys are typically calibrated in terms of either field strength, field strength squared units, or plane wave equivalent power density. As will be seen, many instruments use detectors that are actually responsive to the square of the field strength, either electric or magnetic, but provide meter readings in plane wave equivalent power density; the meter indications are obtained via an electronic manipulation equation. In other cases, instruments designed to sense the magnetic field are calibrated in terms of plane wave equivalent electric field strength.

Other parameters of importance are the frequency or frequencies of the principal source(s) and the duration of the exposure. Virtually all current standards for electromagnetic field exposures are frequency dependent; this dependence takes into account the frequency response of the human body in terms of energy extraction from the incident fields. Because the body acts much like a radio antenna absorbing energy from the field better at certain frequencies, most recommended limits for exposure levels vary in accordance with this frequency selectivity of the body, with the most stringent controls in the body resonance range and less restrictive levels at frequencies both lower and higher than this range. Therefore, when performing a survey of exposure, knowledge of the frequency of the field is necessary to relate to permitted exposure limits and the instrument being used for the measurements must be capable of accurate response at the frequency of exposure.

Electromagnetic fields have a wide range of waveforms depending on their use. For example, a pulsed radar signal typically will exhibit a very high peak to average field strength ratio due to the use of very narrow pulse widths but very high peak powers. In the case of normal AM radio broadcasting, again the average value of the field will be dependent on the programming material and adjustments to the transmitter. TV broadcasting involves the transmission of synchronization pulses which have considerably greater peak powers than the associated video programming material. The waveform of some fields may be of interest because of their rapid changes in time, which, for magnetic fields, may be important from the standpoint of induced currents in the body.

Broadband Instruments

In terms of convenience, broadband instruments are considerably more popular than narrowband instruments for evaluating exposure to electromagnetic fields. Broadband instruments, as their name implies, respond over a wide range of frequencies, possessing nearly flat responses independent of frequency within their frequency pass band. Most common broadband meters feature isotropic probes, consisting of three mutually orthogonal detecting elements connected electronically in such a manner that the output of the

[4] See FCC OET Bulletin No. 65, *Evaluating Compliance with FCC Guidelines for Human Exposure to Radiofrequency Electromagnetic Fields, Supplement A*, August 1997, p. 29.

meter becomes essentially independent of probe orientation within the field. Isotropy is typically achieved to within ±0.5 of 1 dB, but no information on the electrical phase of the fields is retained.

Probe detection elements are normally either thermocouples or diodes. Thermocouples exhibit the useful property of being average-responding detectors. When placed in a modulated field, such as a pulsed radar field, they will accurately respond to the average value of the field squared or average plane-wave equivalent power density. This same property, i.e. true root mean square (RMS) detection, of thermocouples, makes them useful in multiple frequency electromagnetic field environments since they can accurately respond to the sum of the average fields present at the probe. Diode detectors, on the other hand, tend to be peak detectors rather than average detectors unless operated within limited field strength ranges where they can exhibit so-called square law responses, i.e. they will respond correctly to the power density or square of the fields. In some instruments, special circuits are used to correct for the nonsquare law operation of diode detectors. This technique works well for single frequencies, but cannot properly correct readings when multiple signals are being measured. In summary, while diode detectors have certain disadvantages, they also have several advantages over thermocouples, including durability, ability to withstand overloads from strong fields, significantly better thermal stability, and high sensitivity.

While isotropic survey instruments have separate probe assemblies connected by cable to the meter, some survey instruments have single or multiple sensing probes mounted directly on the instrument package. Those instruments equipped with one probe element are responsive to only one field polarization component at a time. The absence of a connecting cable between the probe and the readout package, in some cases, is responsible for superior performance when used in low-frequency fields; it is common for cables to "pick up" or couple with the electric field at lower frequencies, typically below about 3 MHz, often leading to erroneous indications of field strength. Cableless instruments, or ones equipped with nonconductive, fiber optic cables, are preferred for electric field measurements at low frequencies.

Broadband instruments can exhibit flat frequency responses over large frequency ranges. For example, electric field probes can be purchased with specified passbands as wide as 200 kHz to 40 GHz. Magnetic field probes have narrower passband characteristics, such as 0.5 MHz to 300 MHz, and are not generally available for UHF measurements. Care must be used in conducting broadband magnetic field surveys in the presence of UHF broadcast stations, since the loop sensors of the instrument may exhibit high frequency resonances leading to erroneous indications of fields.

The most common broadband isotropic field strength meters, well suited to measurements taken for purposes of ANSI compliance determination, are shown in Table 9.4-4.

Table 9.4-4

Make & Model Series	Probe Freq. Range	Suitable Applications
IFI RFH	E-field: 0.01–220 MHz	AM
Holaday HI-3000	E-field: 0.5–6000 MHz	FM, VHF, UHF, some microwave
	H-field: 0.3–300 MHz	AM, FM, VHF
Narda 8700	E-field: 0.003-40 GHz	FM, VHF, UHF, microwave
	H-field: 0.3–300 MHz	AM, FM, VHF

Most consulting engineering firms own meters of this type to support their work in this field, and some of the meters are available, as well, from certain equipment leasing companies. The total accuracy of these meters is typically ±2 to 3 dB, due to variations in frequency response, isotropicity, and repeatability. Other errors can be introduced, as well, but none which would cause the meter to read low, according to the manufacturers' literature. Therefore, while the measurements are expected to be conservative, leeway should be allowed when establishing the locations of ANSI-threshold contours.

There is also a measurement probe that accounts for the frequency variation of present ANSI RFPG by weighting strength of the fields according to frequency such that it indicates in units of percentage of the exposure criterion permitted. As with any broadband type of instrument, however, it does not indicate how the resultant field magnitude is distributed across the frequency spectrum.

Narrowband Instruments

Broadband instruments, while convenient for conducting area surveys and rapidly assessing exposure levels of electromagnetic fields, do not yield information on the frequency of the fields at a measurement point. Because of the frequency dependence of most RF field protection guides, however, such information is often desired, particularly where frequencies are present for which different exposure limits apply. If total RF power density exceeds 1 $\mu W/cm^2$ in such a situation, the contribution that each field makes to the total must be determined in order to know whether the protection guide is actually being exceeded.

A tunable receiver is the most common instrument for this purpose, allowing each emission frequency to be individually tuned and measured. A spectrum analyzer, which repetitively scans between a lower and upper frequency, producing a graphical display of the measured signal strengths in that range, permits rapid assessment of the relative signal amplitudes across a wide range of frequencies. It is particularly effective when the frequency of a specific source is unknown or when the presence of the field is intermittent.

In either case, however, narrowband meters and their companion sensing antennas are not as portable as broadband meters. It can be very inconvenient or simply impractical to conduct the measurement using the large antennas typically employed with narrowband instruments. This is a serious consideration when performing field measurements under difficult circumstances, such as on tall broadcast antenna towers, and in other situations where it is impractical to carry narrowband equipment. In these cases, arrangements must be made to operate the various sources individually, so that their individual contributions can be assessed with a broadband type of instrument.

Special Adaptations or Applications of Instruments

In the last several years, data processing hardware has been introduced that can be configured with portable RF survey instruments to simplify the data collection task. These devices, which have taken the form of microprocessor based data-loggers or analog devices for simply integrating exposure over some prescribed time interval, now allow direct assessment of the time-averaged exposure to electromagnetic fields. The attractiveness of such devices is that one can use them to actively manage personnel exposures in areas where the exposure levels may reach substantial values momentarily. By constantly monitoring the time-averaged value, the exposed individual has the knowledge to remove himself from the field to keep the average level within acceptable limits.

An area of instrumentation highly developed in the ionizing world of radiation, personal dosimetry, has not been successfully developed for the general case of electromagnetic fields. This has been the case because of the difficulty in relating radio frequency field strengths determined very close to the surface of the body to equivalent, unperturbed values from which SAR may be derived. At lower frequencies, however, personal dosimeters have been developed for use in power frequency fields, i.e. 50 and 60 Hz. At these frequencies, especially for magnetic fields, body perturbation effects can be accounted for in a simpler manner than for higher radio frequencies. Some of these dosimeters contain advanced digital circuitry which allow the retention of data on exposure levels versus time of day. Hence, very complete exposure histories can be determined for subjects wearing these devices. Recently, new developments in the form of personal RF monitors useful in the microwave frequency range have been commercially announced.

Measurement Techniques

There are several different approaches to using broadband exposure meters, depending on the type and number of stations at the site of interest and the purpose of the measurements. In general, a top-down approach is recommended, ensuring that no areas of the site are neglected while also ensuring that potential problem areas are examined in sufficient depth.

The basic technique is to make a quick pass first, throughout the site, of ambient fields, to determine where to concentrate and what fields/boundaries to locate and record. It may be, for instance, that an area exists in which the ambient fields exceed the ANSI RFPG; this is the type of area that would be examined next. There, the location of the ANSI-threshold contour should be defined carefully, for fencing or marking at the time the mitigation measures are determined.

Next, the localized measurements are done, looking for the so-called "hot-spots." The FCC adopted under Docket 88-469 a relaxed distance requirement for measuring hot-spots. In the concluding paragraph (42) of the *Report and Order*, the following statement by the Commission provided somewhat conflicting directions:

"In summary, it is recommended that during routine measurements of radio frequency fields for compliance purposes, a minimum separation distance of 20 cm be maintained between a re-radiating object and the closest sensing element of a probe. However, as a precautionary measure, it is also recommended that consideration be given to the presence of intense, localized fields in the range of 10 to 20 cm from a reradiating object."

This paragraph has been interpreted by some with emphasis on the first sentence, so that measurements are generally made no closer than 20 cm from reradiating objects, but with sufficient measurements made at a distance of 10 cm to determine whether the RFPG is exceeded where "intense, localized fields" are present. Others interpret the paragraph more conservatively, reasoning that, without consistently measuring as close as 10 cm, one cannot give "consideration" to the presence of intense, localized fields in the 10 to 20 cm range. Nevertheless, this increase in measurement spacing, from 5 cm specified in the ANSI RFPG, has alleviated the worries at many sites about localized fields, which tended to occur frequently atop wire-mesh fences or at gates, locations where whole-body SARs would probably remain low but where reproducible measurements indicate fields in excess of the RFPG.

Finally, measurements are made inside the transmitter buildings at the site, looking for both ambient and localized fields and searching with particular care, since these are generally areas of prolonged exposure. Additional grounding will often remove or reduce the severity of localized fields. It is also often helpful for later reference, when determining mitigation measures, to temporarily ground possible sources of reradiation to evaluate the effectiveness of grounding in that specific case.

APPENDIX

Further Guidance for Broadcasters Regarding Radio Frequency Radiation and the Environment

(Federal Communications Commission Public Notice, January 28, 1986)

The National Environmental Policy Act of 1969 (NEPA), 42 U.S.C. Sections 4321–4361, requires all federal agencies to ensure that the environment is given appropriate consideration in agency decision-making.

In a *Report and Order* in General Docket No. 79-144, 100 FCC 2d 543 (1985), the Commission decided that human exposure to radiofrequency (RF) radiation was a proper environmental concern of this agency and specified that the guideline for determining the significance of such exposure will be the "Radio Frequency protection Guides" adopted in 1982 by the American National Standards Institute (ANSI C95.1-1982). As of January 1, 1986, all applications for new facilities, modifications to existing facilities, and renewals must contain *either* a specific indication that the RF radiation of the particular facility or operation will not have a significant environmental impact *or* an environmental assessment which will serve as the basis for further Commission action.[1] *See* Part 1, Subpart I of the Commission's rules for specific regulations regarding environmental matters.

Most broadcasting facilities produce high RF radiation levels at one or more locations near their antennas. That, in itself, does not mean that the facilities significantly affect the quality of the human environment. Each situation must be examined separately to decide whether humans are or could be exposed to high RF radiation. Paragraph 37 of the *Report and Order* points out that accessibility is a key factor in making such a determination. As a general principle, if areas of high RF radiation levels are publicly marked and if access to such areas is impeded or highly improbable (remoteness and natural barriers may be pertinent) then it may be presumed that the facilities producing the RF radiation do *not* significantly affect the quality of the human environment and do not require the filing of an environmental assessment.

Because we wish to avoid burdening applicants with unnecessary work, expenses and administrative filings, we offer the following guidance as to how we will view typical situations. The term "high RF level" means an intensity of RF radiation, whether from single or multiple sources, which exceeds the ANSI guidelines.

Situations

A. High RF levels are produced at one or more locations above ground level on an applicant's tower

If the tower is marked by appropriate warning signs, the applicant may assume that there is no significant effect on the human environment with regard to exposure of the general public.

B. High RF levels are produced at ground level in a remote area not likely to be visited by the public

If the area of concern is marked by appropriate warning signs, an applicant may assume that there is not significant effect on the human environment with regard to exposure of the general public. It is recommended that fences also be used where feasible.

C. High RF levels are produced at ground level in an area which could reasonably be expected to be used by the public (including trespassers)

If the are of concern is fenced *and* marked by appropriate warning signs, an applicant may assume that there is not significant effect on the human environment with regard to exposure of the general public.

D. High RF levels are produced at ground level in an area which is used or is likely to be used by people and to which the applicant cannot or does not restrict access

The applicant must submit an environmental assessment. This situation may require a modification of the facilities to reduce exposure or could lead to a denial of the application.

E. High RF levels are produced in occupied structures, on balconies, or on rooftops used for recreational or commercial purposes

The applicant must submit an environmental assessment. The circumstances may require a modification of the broadcasting facility to reduce exposure or could lead to a denial of the application.

F. High RF levels are produced in offices, studios, workshops, parking lots, or other areas used regularly by station employees

The applicant must submit an environmental assessment. The circumstances may require a modification of the facilities to reduce exposure or the application may be denied. This situation is essentially the same as E. We have included it to emphasize the point that station employees as well as the general public must be protected from high RF levels. Legal releases signed by employees willing to accept high exposure levels are not acceptable and may not be used in lieu of corrective measures.

G. High RF levels are produced in areas where intermittent maintenance and repair work must be performed by station employees or others

ANSI guidelines also apply to workers engaged in maintenance and repair. As long as these workers will be protected from exposure to levels exceeding ANSI guidelines, no environmental assessment is needed. Unless requested by the Commission, information about the manner in which such activities are protected need not be filed. If protection is not to be provided, the applicant must submit an environmental assessment. The circumstances may require corrective action to reduce exposure or the application may be denied. Legal releases signed by workers willing to accept high exposure levels are not acceptable and may not be used in lieu of corrective measures.

The foregoing also applies to high RF levels created in whole or in part by reradiation.

A convenient rule to apply to all situations involving RF radiation is the following:

[1] In applications for new and modified facilities, the requirement for a specific indication is satisfied by answering the question on the form regarding environmental matters. An environmental assessment is the narrative statement described in Section 1.1311 and elsewhere in the Commission's rules.

1. Do not create high RF levels where people are or could reasonably be expected to be present *and*
2. Prevent people from entering areas in which high RF levels are necessarily present.

Fencing and warning signs may be sufficient in many cases to protect the general public. Unusual circumstances, the presence of multiple sources of radiation, and operational needs will require more elaborate measures.

Intermittent reductions in power, increased antenna heights, modified antenna radiation patterns, site changes, or some combination of these may be necessary, depending on the particular situation.

For further discussion *see* Office of Science and Technology Bulletin No. 65, "Evaluating Compliance with FCC-Specified Guidelines for Human Exposure to Radiofrequency Radiation," October 1985. Copies of this bulletin may be ordered from the National Technical Information Service, (800) 336-4700, Order No. PB 86-127081.*

For further information regarding this *Notice*, applicants should contact the AM Branch, FM Branch, TV Branch, or LPTV/Translator Branch, as appropriate.

* NOTE: A limited number of copies are also available from the FCC, (202) 653-8169.

9.5
ELECTRICAL SHOCK

CLIFFORD D. FERRIS
UNIVERSITY OF WYOMING, LARAMIE, WY

INTRODUCTION

The possibility of electric shock must be taken very seriously in the operation of electrical, electronic and broadcast systems because it can produce permanent injury and death. The best approach to electric shock is prevention. Human tissues individually, and the human body as a whole, are electrical conductors. An electric shock is produced when a potential difference is applied across human tissues such that a response current results. The severity of the resulting shock is a function of skin quality, contact pressure, nature of the voltage source (ac or dc), and frequency when alternating current sources are involved.

General Definitions

Human body tissues, both solid and fluid, are classified as *electrolytes*. Thus they are electrically conductive and may be characterized at dc and low frequencies as ohmic resistances. As frequency increases, human tissues also manifest capacitive properties. This latter attribute is of considerable importance with respect to broadcast and communications equipment.

Two general categories of electric shock are recognized: *microshock* and *macroshock*. The former describes an inadvertent electric shock that occurs within the body, and associated electric current levels range from 10 μA to 100 μA. Such a shock may result from improper electrical grounding techniques associated with various surgical and diagnostic medical procedures (such as cardiac catheterization) in which electrically operated sensors are placed internally in the body.

Macroshock (electric shock) is the result of simultaneous physical contact between the external body surface and two or more electrical conductors at different potentials. Electric shock can be induced by body contact between energized conductors, such as two of the phase conductors in a three-phase system, or by contacting a hot wire and neutral or ground.

Current Levels

The electric current levels and associated physiological responses are somewhat different between ac and dc. Table 9.5-1 presents typical data for alternating current shocks at 60 Hz (50 Hz data are similar).

Generally speaking, women react to electric shock at current levels that are about two-thirds of those necessary to elicit the same response in men. For example, current level versus frequency studies published by C.F. Daziel indicate that the mean *let go threshold* at 60 Hz for women is approximately 10 mA as opposed to 16 mA for men.

Table 9.5-1
60 Hz Electric Shock Current RMS Levels and Physiological Responses

Current Level	Physiological Response
1 mA	Lower limit for sensation of shock
5 mA	Upper limit for what is normally considered a harmless shock
10–20 mA	Let go threshold: contractions of *flexor muscles* are stronger than extensions of *extensor muscles*. Victim cannot let go of conductor and begins to perspire
30–40 mA	*Tetany* with sustained contraction and cramping of muscles
50–70 mA	Intense pain with physical exhaustion, fainting, permanent nerve damage, possible respiratory arrest and asphyxiation; potential for ventricular fibrillation (heart)
100 mA	Death by cardiac arrest if the current passes through the body trunk
>100 mA	*Ventricular fibrillation;* if shock is survived: severe burns, amnesia, severe electrolysis at body contact sites
>5 A	Little chance for survival

By a strange quirk of nature, the electrical impedance of the human body is at a minimum between approximately 30 Hz and 80 Hz. This situation, along with other factors to be addressed subsequently, renders a sustained electrical shock at 60 Hz and 120 VAC rms potentially extremely lethal. The threshold levels indicated in the table increase markedly on either side of the low impedance frequency window. The average value for the let go current at dc is on the order of 75 mA. This value is greater than the corresponding ac peak-to-peak value at the 20 mA rms level. The dc threshold of sensation level is increased to 5 mA. On the high-frequency side of the window, the let go threshold increases to an average value of approximately 60 mA at 5 kHz. A momentary direct shock current path through the heart of 60 mA rms at 60 Hz can induce *ventricular fibrillation,* while 300 to 500 mA is required for dc.

Primary and Secondary Effects

There are both primary and secondary effects associated with electric shock. The primary or direct effect is the passage of an electric current through body tissues. Secondary effects include immediate physiological damage, psychological shock, physiological shock and involuntary muscle contractions. In some cases, the latter are strong enough to cause the victim to be thrown some distance from the source of the shock, which can then result in broken bones and/or death if massive objects are impacted.

Various factors control the severity of an electric shock. The intensity of the electric current that is driven through the body as a consequence of the externally applied potential difference is the cause of the reversible and irreversible effects produced by an electric shock. The electrical resistance of the body volume involved and the contact pressure and surface area at the electrical contact sites have major influence on the severity of a shock. The potential seriousness of an electric shock is influenced by the paths that the shock current takes through the body. Current paths that do not involve the body trunk or head are normally not fatal, although permanent injury may result in the form of partial or complete paralysis of a limb, neuromuscular damage, loss of limbs and scars. A severe electric shock to the head may produce loss of memory, loss of motor function (voluntary muscle motion) or death. Shock current paths that pass through the body trunk in the region of either the heart or lungs can cause death. A current path through the heart may induce cardiac arrest, while a path through the lungs can paralyze the nerves that control the breathing diaphragm. The two most dangerous current paths through the chest are from hand-to-hand and hand-to-leg or foot, especially the path from the left hand to the left leg/foot (see Figure 9.5-1).

Human response to an electric shock is both objective and subjective. The condition of the skin surface that comes into contact with the shock source and the contact pressure are major factors, since they determine the contact interface resistance. If an adult tightly grasps the probes of an ohmmeter between the thumb and forefinger of each hand, the resistance that is subsequently measured will vary widely across individuals. Individuals whose hands are calloused from physical activity may generate resistance readings on the order of 1 or 2 MΩ with dry hands, and on the order of 100 kΩ when the hands are wet. Individuals with soft and pliant skin may generate readings of 5–10 kΩ with dry hands, and as low as a few hundred ohms when the hands are wet or damp with perspiration. The resulting current changes by a factor of 1000 from a 1 MΩ contact resistance to a 10 kΩ contact resistance.

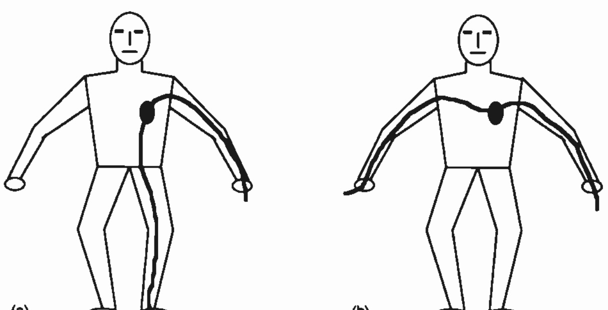


Figure 9.5-1. Most lethal electric shock paths: (a) left hand-to-left foot; (b) hand-to-hand.

Individuals who experience the same low level electric shock may experience very different sensations. Among the sensations reported are tingling (pins and needles), burning, buzzing and jolts. To a large extent, the body location of the shock and the emotional state of the individual at the time may strongly influence response. Response to an electric shock tends to be amplified in individuals who are already tense. Shocks that occur close to major nerves or nerve endings that lie close to the skin surface also tend to elicit amplified reactions. An analogous example is the reaction when one bumps an elbow and triggers the funny-bone response.

There are several secondary effects of electric shock that may produce serious injury even when the shock itself is not life threatening. Most commonly the startle reflex is triggered, which may cause the victim to flail limbs, fall down, drop objects or otherwise move in a manner that causes injury. High voltages, including those produced by capacitor discharges, can produce generalized severe muscle contractions of sufficient intensity to throw a person across a room. Because of the violence of such reactions, the victim may sustain serious bone fractures, internal injuries or even death from impact with nearby objects. The violence of the muscle contractions alone may be sufficient to fracture spinal vertebrae, produce shoulder dislocations and other injuries.

DIRECT PHYSIOLOGICAL EFFECTS

An examination of Table 9.5.1 indicates that the medical consequences of an electric shock range in severity from no effect at all, to minor burns, muscle and nerve damage, and in the extreme case, death. Factors that affect the physiological consequences of electric shock are:

- The effective electrical resistance between the shock contact points on the body
- The portion of the body traversed by the shock current path(s)
- Type of electrical source that produced the shock
- Victim's body weight.

The first two items have been addressed to some extent in the introductory section. As contact resistance at the shock points decreases, the resulting shock current increases along with the potential for severe physiological damage. Current paths through the head and the body trunk are potentially the most lethal. The electrical properties of the human body are frequency dependent such that minimum electrical impedance and consequently maximum shock currents occur in the frequency range from 30 to 80 Hz. Shock threshold parameters are lower for women than men by approximately 30%, and may result from differences in skin condition and associated electrical resistance.

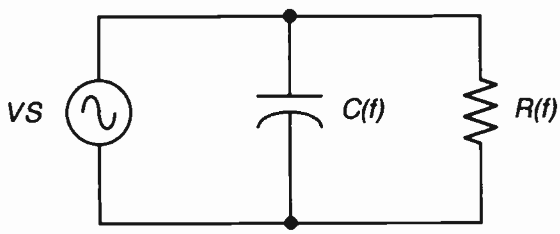


Figure 9.5-2. Elementary electrical model for tissue impedance.

Shocks that exceed the let go threshold are particularly dangerous because the victim begins to perspire. The resulting increase in perspiration moisture on the skin surface reduces contact resistance at the shock sites, thus increasing the shock current and the effects thereof.

Alternating current shocks at 50–60 Hz are potentially more serious than dc shocks because the body impedance has a minimum value. Radio frequency shocks have some additional attributes that will be discussed subsequently.

Soft Tissue and Skeletal Damage from Electric Shock

Human tissues are electrolytes and they can be grossly characterized electrically by the simplistic parallel resistance-capacitance model illustrated in Figure 9.5-2 (*VS* represents the applied electric shock voltage). Both the resistance and capacitive components of tissue are functions of frequency, which is why the model elements are shown as R(f) and C(f). Only the resistive element is significant at dc. As frequency increases into the radio frequency range, both the resistance and capacitance values decrease. The relative change in these parameters results in the minimum value of body impedance in the 30–80 Hz range. At radio frequencies, the capacitive component of the current is significant.

Electrolysis is a primary effect of the passage of an electric current through tissue. Ionic dissociation of the electrolyte (tissue) occurs with the production of heat. When current level is high, steam is generated from the tissue water component, with the consequent rupture of cell membranes and cell destruction. Sustained passage of current causes coagulation of proteins and tissue death along the current path. Medical effects include surface burns at the shock contact sites, penetrating burns along the current path (especially when radio frequency currents are involved) and the potential for permanent damage to both nerve and muscle tissues.

Traditionally, thermal burn severity has been classified as first, second or third degree, with a designation of fourth degree for certain burns resulting from electric shock. Thermal burns resulting from localized contact with a hot object differ from those produced by electric shock in that only one circumscribed portion of the body surface is affected. With electric shock, burns may occur at all contact sites. Burns may be categorized as follows:

- **First-Degree.** Superficial with local discoloration or reddening of the skin in the burn area, mild swelling and associated pain. Healing is normally rapid and without medical complications unless the victim has some underlying medical condition such as diabetes
- **Second-Degree.** Burn damage penetrates more deeply into the skin than in the first-degree case. The skin at the burn site develops a red or mottled aspect with blister formation and substantial pain. The area may swell considerably over the course of several days. The skin surface becomes wet from loss of plasma from the damaged skin layers. Medical attention should be sought to prevent secondary infection, and especially in cases of underlying medical disorders
- **Third-Degree.** Deep penetration into the skin occurs, often into the underlying tissue, with coagulation of the skin tissues, destruction of red blood cells and charring of the skin. The skin surface may initially appear white or charred, or it may resemble a second-degree burn. Full healing of the skin may occur only at the edges of the burned area, with scar tissue replacing the normal tissue in the central portion of the area injured. Risk of infection at the burn site is high and medical attention is important
- **Fourth-Degree** (electrical burns). Overlying tissues are charred such that underlying bone is exposed. This type of injury can be sustained from contact with high-voltage transmission lines or the high-power output section of a radio frequency transmitter. If the incident does not result in death, permanent physical disability can be anticipated
- **High-frequency.** Deep penetration of the tissues at the contact sites. Intense electrolysis and searing of the tissue. Fourth-degree burns may result. The shock current may pass over the skin surface (because of the capacitive component of the skin tissue) as well as through a portion of the body. Multiple damage sites may occur, especially when the current passes over the skin surface and then arcs to grounded objects in the near vicinity. Because high-frequency burns are deep with severe electrolysis, they are very slow to heal, and very painful. Immediate medical attention is important and necessary.

Burns produced by high-frequency energy have beneficial medical uses although they may have grave consequences in the work place. *Electrosurgery apparatus* (ESA), also called an *electric scalpel*, is used to provide the surgical functions of cutting, cauterizing or simultaneous cutting and cauterizing. In effect, a controlled high-frequency shock current is applied at the surgical site (shock entrance site). A large-contact-area return-path (exit) electrode is located elsewhere on the body. The current density at the return electrode is normally very low so that skin injury does not occur. ESA apparatus typically operates in the 300–400 kHz range. Appropriate modulation of the waveform produces the surgical functions required. Even when ESA is used under highly controlled conditions, extraneous burns may occur to the patient if the surface conduction component of the current finds an alternative path to ground (such as an area of the skin surface that has accidentally contacted some grounded metal object that is present in the operating room).

Electric arcs are another source of electrical burns, which may or may not be associated with electric shocks. Burns can be produced by heat and flying hot debris from electric arcs even if the victim is not in electrical contact with the associated circuit. The temperatures within an electric arc can be on the order of 5000° C and above. Fourth-degree burns may be expected if an arc occurs in conjunction with and electric shock.

Skeletal damage may be a secondary effect produced by an electric shock. High voltages, either from momentary contact with a continuous source of electrical energy or a charged capacitor, generally cause violent muscle contractions. Such muscle contractions may be of sufficient intensity to break bones, or cause joint dislocations. In some cases, the victim is propelled away for the shock source. Broken necks and other serious skeletal injuries have been reported when the victim has been thrown against some object (such as a table or workbench).

Shock Paths—Limbs and Extremities

In most cases, electric current paths restricted to the limbs and extremities (excluding the head) are not fatal, although permanent damage and/or disability may be consequences. Loss of a limb may occur, and permanent skin scars, partial or total limb paralysis, and physical impairment are not uncommon.

Shock Paths—Head

High intensity shocks to the head may cause permanent injury to the neck and spine (as a result of body spasms), loss of memory and/or motor function and permanent brain damage. Various bones in the body may be fractured as a result of severe and generalized muscle contractions. At one time electroshock therapy, in which high intensity shock currents were passed through the temples, was used medically to treat patients with certain mental disorders. There are records in the medical literature of the many side effects of such treatments.

Shock Paths—Body Trunk

When the shock current passes through the heart or lungs in the body trunk, grave consequences may occur. The human body is a complex electrical machine. Sensory information from our environment is translated into electrical pulses that are transmitted along nerve pathways to sites in the brain for central processing. Our senses of sight, smell, hearing, temperature and touch depend upon electrical pulses. Our voluntary actions associated with muscle movement and locomotion, as well as involuntary actions such as digestion, heartbeat and respiration, all depend upon electrical pulses. Muscle motion is activated when pulses are transmitted along nerves to *myoneural junctions* where they then initiate muscle extension or contraction.

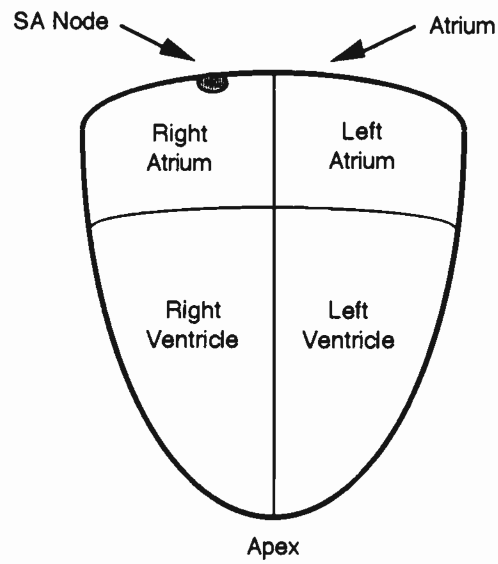


Figure 9.5-3. Stylized anatomical cross-section of the heart.

The heart mechanically is a simple four-chambered pump that is composed of muscle tissue. Its pumping action is controlled by electrical signals generated by a physiological multivibrator, called the *pacemaker* or *sinoatrial (SA) node*. The SA node is a small region of specialized tissue located at the top of the heart. A stylized diagram of the heart is presented in Figure 9.5-3.

The right and left atria that form the two upper chambers of the heart are reservoirs that store the blood that is returned to the heart from the lungs and the body organs. Through appropriate valves, these atrial chambers supply blood to the ventricles, which are the two lower and larger pump chambers. When the ventricular muscles contract to produce a heartbeat, blood from the right ventricle is pumped through the pulmonary artery to the lungs (the pulmonary circulation); blood from the left ventricle is pumped through the aorta for distribution to the body organs (the systemic circulation). Additional valves in the ventricles prevent blood backflow. The beating of the heart is controlled by the regular electrical pulses that are generated by the pacemaker, which cause the ventricles to contract in unison (one pulse per heartbeat). The repetition frequency of the pacemaker in a normal heart automatically adjusts to the body's demand for oxygen through a complex physiological feedback system.

Contraction and relaxation of the skeletal muscles also generates electrical pulses. Contraction of the heart is initiated by electrical pulses generated by the pacemaker, but at the same time, the contraction and relaxation of the heart muscle tissue produces electrical pulses. This electrical activity caused by the beating of the heart can be detected on the external body surface because the human body is an electrical volume con-

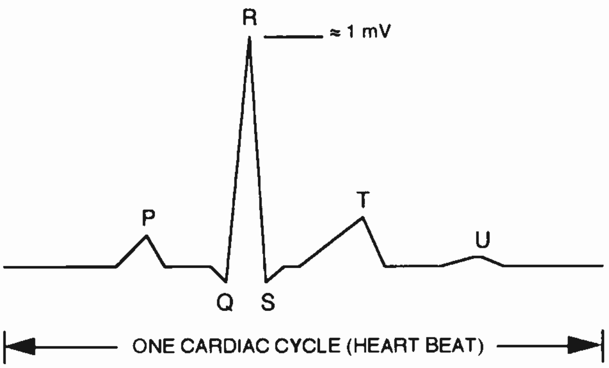


Figure 9.5-4. Stylized ECG waveform for one cardiac cycle.

ductor. The electrical impulses so produced can be detected and recorded electronically to produce an *electrocardiogram* (ECG). There are several protocols and body locations for placing electrodes on the surface of the body to record an ECG. When a reference electrode is placed on the inside of the right ankle and two recording (active) electrodes are placed respectively on the insides of the left ankle and left wrist, a signal of the form illustrated in Figure 9.5-4 can be obtained. It is a representation of the Lead II signal produced by one of the standard methods for recording the electrical activity from the heart. The presentation has been linearized and stylized for purposes of discussion. Actual ECG records are somewhat less linear and the peaks of some of the pulses are rounded. The illustration represents one cardiac cycle or heartbeat. In the Figure, the time axis has been normalized to a length of one heartbeat. The major events in the record (waves) are designated as P, Q, R, S, T, and U. The peak voltage amplitude produced by the R wave on the body surface typically ranges from approximately 1 mV to 1.5 mV. These waves or events relate to the beating of the heart as follows:

- The P wave represents the electrical activity produced when the atria contract to force blood into the ventricles
- The R component of the QRS complex represents the contraction of the ventricles and the pumping of blood out of the heart and into the aorta and pulmonary artery
- The T wave represents the electrical activity of the relaxation (called repolarization) of the ventricular muscles following their contraction
- The electrical signal produced by the relaxation (repolarization) of the two atria is masked by the high amplitude R wave
- The origin of the U wave is somewhat unclear, but is thought to be produced by the repolarization of the aorta after the blood volume produced by the left ventricle has passed through its length.

When an intense electrical shock current passes through the chest cavity, it may cause the entire heart muscle to contract and cease beating during the time duration of the current. If this shock current is of short duration, normal heart rhythm usually returns following the passage of the current and there is no irreversible cardiac tissue damage. Less intense shock current levels can momentarily disrupt the regularity of the normal heart rhythm. The heart is most sensitive to external electrical stimulation during the occurrence of the T wave. A shock current amplitude that does not ordinarily disrupt cardiac rhythm during other portions of the cardiac cycle may produce cardiac arrest if it occurs during the T wave event.

Either of two conditions is described under the designation cardiac arrest: *cardiac standstill* and *ventricular fibrillation (VFIB).* As the name implies, the heart does not beat in the condition of standstill, and the ECG is a flat line. Ventricular fibrillation describes a condition in which individual muscle groups of the ventricles contract randomly and out of unison. No blood is pumped and the blood pressure drops to zero, but a low level random ECG signal is produced, as is shown in Figure 9.5-5. To differentiate between the conditions of flat line and VFIB, an ECG recorder is required.

Involuntary contraction of the respiratory muscles is often a consequence of the passage of an electric shock current through the chest region. Since respiration ceases, the victim can die from asphyxiation if the current is not interrupted. Cessation of breathing (respiratory paralysis) can occur if the electric shock current that passes through the body trunk damages the nerve that controls the diaphragm. In such a case, the victim typically inhales deeply; the chest then expands and becomes rigid and remains so after the shock current has ceased. Initially, the heart may or may not be affected. In the latter case, the heart continues to beat, but the beats become irregular and erratic with time as the blood carbon dioxide level increases (acidosis). Acidosis produces irritability of the heart muscle tissue and makes restoration of a normal heartbeat difficult even if respiration can be restored.

Dangers of 120 VAC rms 60 Hz Electric Shocks

The electrical impedance of the human body is at a minimum value for 60 Hz. A voltage level of 120 VAC rms is generally not sufficiently high to cause muscle contractions that throw the victim away from the source of the shock current. It is also low enough that a strong startle reflex may not occur. Consequently

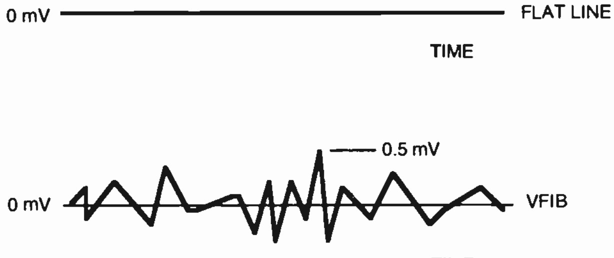


Figure 9.5-5. Cardiac arrest ECG waveforms.

the subject may not be able to let go of the source if the conductor has been firmly grasped. If the effective impedance presented by the body to the shock source is 500 Ω, then the shock current will be 240 mA rms, which is more than enough to produce ventricular fibrillation. Even if a higher impedance was presented initially, the continuation of the shock produces perspiration buildup on the skin that lowers the effective resistance. Shock charges on the order of 100 mA seconds (rms current values) are sufficient to produce VFIB. The cardiac cycle at a heart rate of 70 beats per minute is 857 ms or 0.857 seconds. At a current level of of 240 mA, the shock energy generated during one cardiac cycle is 206 mA seconds. Thus VFIB can be anticipated as one of the consequences of a sustained 120 V 60 Hz shock.

ELECTRICAL SHOCK PREVENTION

Because death or permanent disability may be a consequence, electric shock must not be taken lightly. Shock prevention is far superior to medical remediation after the fact. Some general guidelines are:

- **Workplace safety regulations.** Observe all company and local regulations regarding electric shock prevention. In many industrial sitings, the regulations promulgated by the Occupational Safety and Health Administration (OHSA) apply as well
- **Clothing and jewelry.** When working in close proximity to exposed energized electrical circuits, loose clothing and jewelry should not be worn (remove neckties, scarves, etc). Exposed skin should be minimized (do not wear shorts, sandals or open-toed shoes, etc.). It is strongly recommended that ALL jewelry (including wrist watches) be removed
- **Protective clothing.** Observe all company regulations regarding the use of safety clothing and accessories. Safety goggles, non-metal hard hats, electrically insulating safety shoes and various safety garments (electrically insulating gloves, for example) may be required. OHSA regulations generally require the use of goggles, hard hats, and safety shoes in most industrial venues
- **Equipment servicing.** Power switches should always be turned off and electrical/electronic equipment should be disconnected from the power mains before opening the equipment cabinet for servicing or calibration of internal circuits. Never rely on switches to disconnect power. Sometimes they are defective, and sometimes a switch may be miswired or mislabeled such that the normal off and on positions are reversed
- **High-voltage capacitors.** High-voltage capacitors associated with power supplies, cathode ray tubes, radio transmitters and other high voltage electronic and radio frequency equipment must be discharged prior to equipment repair or servicing. Use appropriate discharge cables designed for this application. Always discharge the external surfaces of cathode ray tubes and other devices on which high static charges can develop.
- **Live circuits.** When diagnostic tests or calibration procedures must be conducted on energized equipment, use all appropriate electrical safety procedures. A large body of modern electronic equipment utilizes switching/regulating power supplies that are potentially lethal because portions of the internal printed circuit boards are hot at the ac power line voltage. For servicing, an isolation transformer must be used between the power mains and the equipment to reduce electric shock hazard. The hazard results from a bridge rectifier that is connected directly across the ac power mains when the equipment is energized. This condition causes the circuit board ground to be at a different potential from the ac power line neutral or Earth ground. Without the use of an isolation transformer, accidental contact between Earth ground and the circuit board ground may result in a lethal electric shock. As general practice, an individual should not work alone on energized circuits. There should always be at least one noninvolved observer who can disconnect electric power and obtain help as required in the event of an accident
- **Earth ground connections.** Never disable or remove the Earth ground connections on power cords and within equipment
- **Equipment design and modification.** All appropriate codes and guidelines should be observed, such as those promulgated by: the National Electrical Code of the National Fire Protection Association; Medical Devices Safety Act of 1971, 1973, and subsequent amendments; American National Standards Institute (ANSI); Institute of Electrical and Electronics Engineers (IEEE); American Society of Mechanical Engineers (ASME); Underwriters Laboratories (UL); Canadian Safety Association (CSA); industry standards for equipment design and manufacture; other regulations and guidelines as appropriate
- **Antenna installation.** Erection of antenna towers and antenna installation requires skilled workers. Mechanical assembly should follow good engineering practices and be in compliance with all local and federal codes. Care must be taken that overhead power lines and other electrical distribution systems are not accidentally touched by personnel, equipment or structural components. Where appropriate, structures must be correctly grounded
- **Transmitters.** The high-voltage power supplies and high-voltage radio frequency tuned circuits associated with transmitters present strong potential for electric shock. Extreme care must be taken by personnel who work in close proximity to such installations. Corona discharge to electrically conducting objects (including human flesh) that come in close proximity to energized circuits is an ever present possibility. Areas of high energy concentration, such as low radius bends in conductors, connections to high-voltage capacitors and inductors, and the like, are likely sources for corona discharge. The potential danger for both electric shock and severe radio frequency burns cannot be ignored.

High-Frequency Hazards

There are other electrical hazards in addition to direct electric shock that are associated with high-frequency equipment. It has been widely assumed that the human body is transparent to electromagnetic radiation below microwave frequencies, although there has been some discussion in the media about 900 MHz cellular telephones and human exposure to low-frequency electric/magnetic fields. Since the 1950s, it has been known that high power microwave transmitters can present extreme health hazards for any personnel who place themselves in line with the transmitting horn or antenna array. Local or generalized elevation in body temperature produced by the absorbed microwave energy can damage internal organs and produce cataracts. This phenomenon, of course, is the basis for microwave ovens.

Ungrounded conducting surfaces located in high flux electromagnetic fields can develop large induced circulating currents. When high power radar and communications transmitters are positioned in close proximity, such as on some ships, an electric discharge hazard can develop. This situation is the basis for the requirement that aircraft be electrically grounded before fueling, or arming in the case of military aircraft, and that radio transmitters be turned off in the vicinity of blasting operations. These induced circulating currents may be of sufficient intensity to activate ordinance if a discharge path is generated, and to produce serious electrical shocks in personnel. Caution should be observed when working close to energized transmitting antennas, regardless of the frequency. Nearby ungrounded conductors may be at dangerous potentials.

FIRST AID

Prevention is the best first aid for electric shock. When a person has sustained or is sustaining an electric shock, several suggested guidelines for remedial action are:

- **Shock in progress.** In the situation when an individual is sustaining an electric shock and cannot let go of the electrical source, the safest action is to disconnect the circuit. This should be accomplished by tripping a circuit breaker in the power mains supply, or by pulling the power line plug that feeds the equipment involved if this can be accomplished safely. Under no condition should the rescuer touch the victim, since the rescuer's body also may then be placed in the dangerous current path. In the event that neither the circuit breaker nor the power plug can be located, then an attempt should be made to separate the victim from the electrical source by physical means using an insulating object. The implement employed should not contain metal or other electrically conducting materials. The optimum device is a *hot stick* (with the use of insulating lineman's gloves) to pull or push the victim away from the hazard. Pulling on a hot stick normally provides more motion control than pushing. Lacking a hot stick, a wooden broom handle or wooden chair/stool may be used in an attempt to free the victim. The rescuer must use extreme caution not to touch either the victim or the shock source, and thus also become a victim. Once the victim has been completely separated from the shock source (either by disconnecting the power or by physical removal), first aid measures should be initiated immediately
- **Post-shock care with conscious victim.** In situations when the victim is conscious and moving about, convince the victim to sit or lie down. Often there is a delayed action to electric shock that causes the individual to faint or collapse. Summon the appropriate in-plant paramedical personnel immediately, or call 911. If the arrival of emergency medical personnel will be delayed, check for electrical burns and any additional immediately obvious injuries. In instances of severe electric shock, there will be burns at both the current entrance and current exit sites. These burns should be covered with (preferably sterile) dry dressings. Examine the victim for possible bone fractures and dislocations if the individual sustained spasms associated with the shock, or was violently thrown from the source (whether or not there was impact with nearby objects). As required, splints should be applied if the rescuer has appropriate training and suitable materials are available. Use a coat or a blanket to cover the victim when the environmental temperature is low, or if the individual complains of feeling cold
- **Post-shock care with unconscious victim.** In the case of an unconscious victim, summon appropriate paramedical personnel immediately. In the interim, check to see if the victim has a pulse and is breathing. Normally a pulse can be felt above the thumb joint at the inside of either wrist (radial pulse), or in the neck above and to either side of the Adam's apple (carotid pulse). It is generally preferable to check the carotid pulse, which is typically stronger and generally easier to locate than the radial pulse. The thumb should not be used to feel a pulse, but rather the index and middle fingers. Many persons have a strong thumb pulse that can be mistaken for the victim's. When a pulse can be felt but the subject is not breathing, mouth-to-mouth resuscitation should be used. If no pulse if felt (and presumably the subject is not breathing), and the rescuer has been trained in cardiopulmonary resuscitation (CPR), then the individual should be moved very carefully to a firm surface and CPR initiated. When the oxygen supply to the brain is cut off, loss of memory or permanent brain damage may occur within a few minutes even when a subject is successfully resuscitated. When a victim has been resuscitated successfully and appears to be stable, follow the previously discussed first aid procedures if there will be a delay in the arrival of paramedical personnel
- **Defibrillator use.** It is ironic that the treatment for cardiac arrest (as a result of electric shock) is a massive countershock delivered from a device called

a *defibrillator*. The massive countershock causes a massive contraction of the heart, and stills the random fluttering produced by ventricular fibrillation when it exists. When conditions are ideal, a normal heartbeat starts spontaneously following the countershock. Portable defibrillator units are available for use by paramedical personnel and other trained individuals. Even when a defibrillator unit is available, it should never be used by an untrained person. If a defibrillating shock is indiscriminately applied to an unconscious person whose heart is beating, cardiac arrest can be induced, just the condition that the instrument was designed to correct.

KEY TERMS

Atrial. Pertaining to the atrium or two reservoir chambers at the top of the heart

Cardiac. Pertains to the heart

Coagulation. The formation or production of a blood clot

Defibrillator. An electronic instrument used in an attempt to shock the heart into beating normally

Electrocardiogram (ECG). A visual record of the heart's electrical activity as recorded from electrodes placed on the body surface; also abbreviated EKG

Electrolyte. A substance that is chemically decomposed by the passage of electric current through it

Extensor muscle. A muscle that extends a joint

Fibrillation. Spontaneous and random contraction of individual muscle groups

Flexor muscle. A muscle that flexes a joint

Pacemaker. A small area of specialized tissue at the top of the heart that produces the electrical pulses that initiate heartbeats

Tetany. The state of a muscle when it is in sustained contraction

Ventricular. Pertaining to the main pump chambers of the heart

BIBLIOGRAPHY

American Academy of Orthopaedic Surgeons, *Emergency Care and Transportation of the Sick and Injured, 3rd. ed.,* George Banta, Menashi, WI, 1981.

American Red Cross, First Aid, "Responding to Emergencies," *Mosby Life Line,* St. Louis, MO, 1991.

Berkow, R., ed., *The Merck Manual of Diagnosis and Therapy, 16th ed.,* Merck, Rahway, NJ, 1992.

Dalziel, C. F., "Electric Shock," *Advances in Biomedical Engineering,* Brown, J. H. U., and Dickson, J. F., eds., volume 3, Academic Press, New York, pp. 223–248, 1973.

Ferris, C.D., *Introduction to Bioinstrumentation,* Humana Press, Clifton, NJ, 1978.

Lee, R.C. et al., eds., "Electrical Injury: A Multidisciplinary Approach to Therapy, Prevention, and Rehabilitation," *Annals of the New York Academy of Sciences, Vol. 720,* The New York Academy of Sciences, New York, 1994.

ADDITIONAL SOURCES OF INFORMATION

In addition to the publications listed under Bibliography, readers can find additional information on human physiology, cardiac function, electric shock and electrical properties of tissue in the following sources:

Ackerman, E., *Biophysical Science,* Prentice-Hall, Englewood Cliffs, NJ, 1962.

Schwan, H.P., "Electrical Properties of Tissues and Cell Suspensions" Lawrence, J.H. and Tobias, C.A., eds., *Advances in Biological and Medical Physics,* Vol. V, Academic Press, pp. 147–209, 1957.

Vander, A.J., et al., *Human Physiology,* 6th ed., McGraw Hill, 1994.

Webster, J. G. ed., *Medical Instrumentation, Application and Design, 2nd ed.,* Houghton Mifflin, Boston, MA, 1992.

9.6

TOWER SAFETY, FALL ARREST SYSTEMS, AND OSHA

LINDA D. KELLEY
CBS CORPORATION, PITTSBURGH, PA

INTRODUCTION

Each year, workplace accidents claim the lives of thousands of workers and cause serious injuries too countless others. Sadly, many of these accidents easily could be avoided. There is no question that it is in the best interest of all employers to reduce exposure to workplace hazards. Besides protecting workers, proactive safety programs save time and money—fewer people off the job as well as lower workers' compensation costs, medical expenses, temporary replacements and salary continuance costs—and they're mandated by law.

The Occupational Safety and Health Administration (OSHA) was created in 1970, as a result of the Act which bears its name. The Act resulted in the promulgation of a series of general and specific workplace standards. Many of those standards have application in the broadcasting arena whether a small radio transmitter site or a large production facility. For example, in the area of life safety, employers are required to have clearly marked and accessible emergency exits, an evacuation plan and some form of fire protection—like sprinklers or fire extinguishers. Electrical hazards, vehicle safety, tripping hazards, asbestos monitoring and maintenance are just a few other examples of safety issues facing many broadcasting locations.

Since OSHA's formation in 1970, numerous regulatory programs have been adopted that apply to worker safety and protection in the broadcast industry.

While this chapter is limited to a review of OSHA requirements related to tower safety (fall protection and lockout/tagout), readers should be aware that there are other OSHA regulations that may apply to broadcasting operations.

Information Resources

OSHA's webpage (http://www.osha.gov) contains links to a wealth of information including the full text of standards, guidelines, technical information and frequently asked questions. (See this chapter's Appendix for details on OSHA's home page.) Also included in the Appendix are the following Code of Federal Regulations (CFR) excerpts:

- 29 CFR Part 1910 Table of Contents to General Industry Standards
- 29 CFR Section 1910.147 Lockout/Tagout requirements
- 29 CFR Section 1910.147/Appendix A Sample Lockout/Tagout Program
- 29 CFR Section 1926.502 Fall Protection Standards
- 29 CFR Part 1926 Subpart M/Appendix C Non-Mandatory Guidelines for Fall Arrest Systems
- 29 CFR Part 1926 Subpart M/Appendix E Sample Fall Protection Plan

Practical Tips

Many tower owners/operators use contractors to repair, inspect or paint their towers. In cases where the owner or operator"s employees do not climb towers, it is recommended that the towers be equipped with appropriate anchorage points only, and that contractors be required to provide their own equipment (harnesses, lanyards, lines etc.).

Owners and operators should then verify that the contractor has the right equipment to use on the tower, that its employees and sub-contractors have been trained and that the contract documents set forth the responsibilities of the parties with respect to safety standards. Be aware, however, that using a contractor or its subcontractors to perform work does not insulate tower owners and operators from the application of OSHA. (See *Case Law* section of this chapter).

Before hiring any contractor, check its safety record either by direct request to the contractor for its compliance history or through OSHA's home page.

When erecting new towers, take fall protection into consideration in the overall construction plan. Include anchorage points, work platforms, guardrails etc., in the design.

Regulatory Requirements[1]

The specific standards applicable to tower safety are fall protection standards, electrical safety standards, and general industry standards. In addition, OSHA imposes an overall "general duty" on employers to provide a safe working environment.

Fall protection applies anywhere a worker is exposed to a fall from an elevation of six feet or more. (1.8 m).[2] The fall protection standards are found at 29

[1]This review covers federal regulations only. Readers are encouraged to check for specific state standards in those jurisdictions where OSHA has approved state programs.

[2]The distance is four feet under 29 CFR 1910.268(g). (See footnote 3).

CFR Section 1926.502.[3] The applicable requirements of this standard are reflected in the fall arrest system criteria, which are reprinted in the Appendix.

The "control of hazardous energy" standards, commonly known as "lockout/tagout," are found at 29 CFR Section 1910.147. These standards apply to exposure to energy sources during servicing and/or maintenance of machines and equipment. The purpose of the standards is simple: to prevent the unexpected energizing or start up of equipment, or release of stored energy, which could cause injury to employees. The full text of the standards are reproduced in the Appendix.

General industry standards—29 CFR Part 1910—cover a wide range of activities. The table of contents to these standards is reproduced in the Appendix. All of these standards, and others, are available through the OSHA website or from the U.S. Government Printing Office.

TOWER SAFETY

The two critical elements of a tower safety program are fall protection and lockout/tagout of energized sources. The law provides that workers must be protected when required to work at elevations of six feet (1.8 m) or more. If engineering-controlled fall prevention mechanisms such as guardrails are not feasible or practicable, a personal fall protection system must be used. In addition, workers must be protected against exposure to energized sources while working on or near a tower.

This section provides an overview of fall protection plans, personal protective equipment and lockout/tagout programs. Relevant case law and enforcement actions also are included.

Guidelines for Establishing a Fall Protection System

These guidelines provide suggested components and strategies for complying with fall protection and lockout/tagout requirements. It should be noted, however, that each situation must be reviewed and evaluated to ensure all potential hazards are addressed and the correct programs are in place to ensure compliance.

Hazard Analysis

Each tower location should be evaluated to determine potential exposure to falls and the type of fall arrest system/personal protective equipment that best addresses each exposure. The evaluation should include a review of all tasks and the mobility necessary to perform those tasks. The analysis should identify any obstructions in the potential fall path and should address escape and/or rescue methods.

Feasibility of Engineering Controls

As noted previously, engineering controls, like guardrails, can be used to address some fall hazards. Towers, however, traditionally require fall arrest systems and personal protective equipment because of fall hazards associated with climbing and descending. Elevated work platforms on some towers (where a worker stands to make repairs, change lights etc.) can be equipped with railings to allow additional mobility and protection while performing those tasks.

The various types of personal protective equipment are described in detail later in this chapter.

Policy and Procedure

Written fall protection/prevention procedures should be an integral part of an employer's overall safety policy. It is not enough, however, to have a written policy on a shelf in an office. Implementation, employee awareness and enforcement of the policy are critical to its success.

A procedure should be developed for each type of potential fall hazard, with details sufficient to address the identified risks. Once appropriate training has been provided and workers are familiarized with fall arrest systems and personal protective equipment, a simple checklist can be developed and reviewed each time the equipment is used.

Written procedures should be reviewed periodically to ensure they reflect changes in the work environment, "lessons learned" etc.

Employee Training

Once a hazard analysis has been performed, written procedures developed and appropriate protection identified all employees who may be exposed to fall hazards must be trained in the procedures and the proper use of equipment.

Training should be reinforced periodically, and all new employees should be trained prior to conducting any work that might expose them to identified risks.

Training Elements

Employee training should address proper methods of equipment use, including capacity limits of equipment; anchoring and tie-off techniques; equipment inspection, storage and maintenance and emergency escape and rescue procedures.

Training aids, such as videos and instructions from equipment manufacturers, can be utilized as a component of training. Some fall protection equipment manufacturers will develop customer-specific videos. In any case, training should be tailored to the various hazard exposures and equipment being used.

[3] Although a specific industry standard exists at 29 CFR 1910.268 for "Telecommunications" (which standard does include some provisions regarding fall protection), OSHA views tower erection, alteration, repair or painting as "construction", covered under 1910.502. Therefore, readers are cautioned not to rely on 1910.268 as the *only* applicable standard to broadcasting operations. (The same is true regarding lockout/tagout of energized sources). See, e.g., 29 CFR 1910.12 (definition of "construction work"); 29 CFR 1910.268 (a)(2) (application of Telecommunications Standard). See, also, OSHA Interpretation Memos and Letters available on this subject on line through the OSHA Home Page or by written request.

In addition, OSHA has published non mandatory guidelines to assist employers in developing training programs. These and other guidelines are available through the OSHA home page and the Government Printing Office. (See the Appendix and Reference for more details).

Program Audit

All employers should conduct periodic audits to ensure that all elements of the safety program are being followed and to determine whether any modifications to the program are necessary.

Lockout/Tagout

Lockout/tagout of energy sources is a critical part of any tower safety program. This includes electric current as well as radiofrequency electromagnetic fields (RFE). Energy sources must be locked out prior to repairs or other work being performed. It is not sufficient merely to shut off electrical power or reduce RFE. The source must be locked out to prevent unknowing re-energization of the source during repairs. For that reason, a locking mechanism must be used. In addition, some form of identifying information—a "tag"—should be used to notify others when, why and by whom the source has been locked out.

Standard locks and tags can be purchased for many types of equipment. In other instances, as with certain transmitters, it may be necessary to have a locking device or hasp designed—either by the manufacturer of the transmitter or an engineer familiar with the operation of the system. It is imperative that workers not be permitted to assume that everyone involved knows that repairs are taking place. Many injuries occur because incorrect assumptions are made.

Because a tower may serve more than one operating entity, it is critical that operators develop a plan to address lockout/tagout and fall protection issues. Besides obvious safety concerns, coordination among the entities is critical to prevent exposure to civil or criminal liability (see the Case Law example).

Consideration should also be given to FCC requirements adopted on August 1, 1996, with respect to occupational exposure to RF electromagnetic fields (RF safety programs, training, warning labels etc.). (See FCC Office of Engineering and Technology (OET) Bulletin 65, Edition 97-01, August 1997 and Supplement A thereto.)

As with fall protection, a hazard analysis, written procedure and training should be performed with respect to energy sources.

FALL ARREST SYSTEMS

As noted previously, a hazard analysis should be conducted at each work location to determine the types of hazards and the best means to eliminate or minimize risk to employees or other workers. This section covers the types of equipment required by OSHA regulations to address climbing and descending hazards.

Personal Fall Protection Equipment

Personal fall protection equipment consists of body support, lanyards and lifelines, restraint lines, connectors, anchorage and tie-offs. The equipment described below meets the elements of a fall arrest system as prescribed by OSHA standards.

Body Support

Body support consists of either a body harness or a body belt. A body harness is a system of straps which, when properly secured about the user, distributes fall arrest forces over at least the thighs, pelvis, waist, chest and shoulders. Body belts, which are secured only around the waist, do not provide the level of security that a body harness gives. OSHA regulations prohibit the use of body belts after January 1, 1998. Keep this in mind when updating an existing fall protection system or developing a new one.

Lanyards and Vertical Lifelines

A lanyard is a flexible line with a connector at each end to connect the body harness to a deceleration device or lifeline. Lifelines can be vertical or horizontal. A vertical lifeline is a flexible line used for connection to an anchorage at one end. Vertical lifelines should be limited to one user per line. Horizontal lifelines are connected between fixed anchor points on the same level. Horizontal lifelines are designed to minimize the potential for pendulum-like swing falls. Horizontal lifelines should be designed, installed and used under the supervision of a qualified person as part of a complete personal fall arrest system, which maintains a safety factor of at least *two.* (A "qualified person" is a professional who can identify hazardous conditions in all aspects of a fall protection program, including application and use of related equipment.)

OSHA regulations require the following with respect to the use of lanyards and lifelines:

- Each employee should have a separate vertical lifeline with a minimum breaking strength of 5,000 lbs (2,268 kg)
- Self-retracting lifelines and lanyards which automatically limit free-fall distance to two ft (.6 m) or less should be capable of sustaining a minimum static tensile load of 3,000 lbs (1,361 kg) applied to the device with the lifeline or lanyard in the fully extended position
- Lifelines and lanyards should not be constructed of natural fibers
- Lanyards should be constructed with at least 0.5 in. (1.27 cm) diameter synthetic rope with a length which limits the opportunity to fall to no more than six ft (1.8 m).

Restraint Lines

Restraint lines may extend from an anchorage or between anchorages to which an employee is secured to prevent a fall from an elevated work surface. Restraint lines should be capable of sustaining a tensile load of at least 3,000 lbs (1,361 kg).

Connectors

Connectors are the devices used to couple parts of the system together, such as "D" rings and snap hooks. A locking snap hook is a connector that requires two separate forces to release it. Locking snap hooks are used to prevent rollout or accidental disengagement.

OSHA regulations require that connectors meet the following criteria:

- Corrosion-resistant finish with smooth surfaces and edges
- Dropped forged, pressed, formed steel or equivalent construction
- Capable of sustaining a minimum tensile load of 5,000 lbs (2,268 kg)
- "D" rings and snap hooks should be tested to a minimum tensile load of 3,600 lbs (1,633 kg) without cracking, breaking or sustaining permanent deformation
- Locking snap hooks should require two separate forces to disengage.

OSHA regulations prohibit the use of non-locking snap hooks as a part of a personal fall arrest system effective January 1, 1998.

Anchorage

An anchorage is a secure point of attachment for lifelines or lanyards that is not a part of the working surface or equipment rigging points. Anchorage points are the key to designing an effective fall protection and arrest system. The strength and utility of the system is based on an anchoring system that will support each of the system's component parts. Therefore, anchorage points should meet the following criteria:

- Capable of supporting at least 5,500 lbs (2,495 kg) per employee using the system
- Designed, installed and used under the supervision of a qualified engineer as part of a complete fall protection system which maintains a safety factor of at least two.

Tie-Offs

A tie-off is the method to connect a user to an anchorage. The following should be considered when placing tie-offs:

- The attachment point to the lifeline or anchor should be located at or above the connection point of the fall arrest equipment to the body harness to ensure a free fall distance of no greater than six feet (1.8 m)
- Obstructions in the potential fall path should be considered when locating a tie-off
- Special provisions need to be made when tying off around an "H" or "I" beam, as the cutting action of the beam edges can reduce system strength by as much as 70%.

(See OSHA guidelines for more details on tie-offs, anchorage, lanyards and other system components.)

Overall System Criteria

A personal fall arrest system should meet the following criteria when stopping a fall:

- Arresting force on user should be limited to 1,800 lbs (816 kg) when used with a body harness
- System should bring user to a complete stop and limit maximum deceleration distance to 3.5 ft (1.1 m)
- System should have sufficient strength to withstand twice the potential impact energy of a user free-falling a distance of six feet (1.8 m) or the free-fall distance permitted by the system, whichever is less
- The attachment point of the body harness should be located in the center of the user's back near shoulder level or above the user's head.

Additional Criteria for Climbing Activities

In addition to the minimum criteria established here, climbing activities require the following system capabilities:

- The user should be able to ascend or descend without the need to continually hold, push or pull any part of the system
- The connection between the lifeline and the point of attachment to the body harness should not exceed nine inches (22.9 cm) in length
- The arrest system should activate within two feet (.6 m) when a fall occurs.

Mountings for rigid carriers should be attached at each end of the carrier. Cable guides used with flexible carriers should be installed no less than 25 ft (7.6 m) and no more than 40 ft (12 m) apart along the length of the carrier.

Ladder safety devices and related support systems should be capable of withstanding a test, without failure, of a 500 lb (227 kg) weight over an 18 in. (45.7 m) drop.

All other personal fall protection systems for climbing activities should be capable of withstanding a test, without failure, of a 250 lb (113 kg) weight over a four foot (1.2 m) drop.

Any fall protection system component that is subjected to impact loading should immediately be removed from service and not reused until it has been inspected and determined by qualified personnel that it is not damaged and is suitable for its intended use.

Inspection and Maintenance

To ensure protection, fall arrest systems should include routine inspection and maintenance. As with other procedures, a written checklist should be developed for use in the inspection and maintenance program. Key elements of the program include:

- System components (safety harnesses, lanyards etc.) should bear identification of the supplier and include the date of manufacture, as well as the OSHA standard to which it was constructed (29 CFR Section 1926.502(e)(5) minimum tensile load for D-rings

and snap hooks). The supplier should certify that the equipment meets the minimum requirements of the appropriate standard
- The manufacturer's instructions should be incorporated into maintenance and inspection (including cleaning) procedures
- Prior to each use, the equipment should be inspected for mildew, wear, damage or other deterioration
- Equipment should be stored in a cool, dry place, protected from UV exposure and the elements
- Equipment should be easily accessible to trained users
- Documented inspections of all protection equipment should occur every six months. (Some manufacturers provide an inspection and recertification service)
- Permanent anchorage points should be inspected periodically by a qualified engineer and immediately following an accidental fall arrest
- Testing should not be permitted when the equipment is in service.

Escape and Rescue

Having a fall protection or lockout/tagout system or procedure is not enough. The law requires employers to address the means by which an employee will be rescued once the exposure occurs. Rescue considerations are part of the OSHA standards (See, for example, 29 CFR Section 1926.502(d)(20) which requires that "the employer shall provide for prompt rescue of employees in the event of a fall or shall assure that employees are able to rescue themselves.") Whether the rescue system is provided directly by the employer, by a contractor or through a local fire department or other emergency team, a rescue plan must be in place in advance of conducting the work.

CONCLUSION

The importance of tower safety programs cannot be overstated. The bottom line is to have a safety program and to enforce it.

RECENT CASE LAW; CIVIL AND CRIMINAL ENFORCEMENT

Employers who violate OSHA standards are subject to civil and/or criminal enforcement. Following are some recent examples of enforcement actions.

Civil Enforcement

An Ohio painting company was cited in February 1997, for willful and serious violations of OSHA requirements by exposing employees to a 183 ft (55.8 m) fall hazard. The citation arose from an accident that left one of the workers paralyzed. Penalties were assessed for $178,000.

On July 2, 1997, it was reported that a Colorado steel company received citations for 61 violations with proposed penalties of $1.1 million (or just over $18,000 per citation). Eight willful citations were issued for items including fall protection and electrical components. Thirty-seven serious violations were identified relating to such things as defective ladders, tripping hazards, blocked exits, electrical components and wiring; 16 repeat violations were included for unguarded floor holes, blocked exits etc.

On August 15, 1997, federal officials levied a fine in excess of $200,000 against a wholesale distributor of doors as a result of a fatal fall suffered by one of the distributor's workers earlier in the year.

The steel erection contractor at Guam International Air Terminal was issued citations and proposed penalties of $8,260,000 for 118 alleged willful violations of fall protection requirements, after a welder fell 65 ft (19.8 m) to his death on March 28, 1995. A settlement was reached in the amount of $1.85 million.

A Texas tower services company was cited for seven serious violations and fined $29,400 for the October 12, 1996 collapse of a TV tower, which killed three men. The family of at least one of the men has sued the company for monetary damages in connection with the incident.

A construction company received six citations and was assessed penalties of $85,000 following an accident in which an employee was electrocuted. The man was killed when the raised boom on his flatbed truck came into contact with overhead power lines.

Another construction company—inspected 16 times in six years—was fined $299,000 for willful violations of fall protection and electric hazard requirements.

Among OSHA's list of the six top egregious cases for 1996 were the following: A Texas bakery was assessed $1,040,000 for willful violations of, among others, lockout-tagout standards; an Indiana steel erector was fined $396,200 after the death of a worker who fell 35 ft; a Boston construction company was assessed $770,000 after the death of one of its subcontractor's employees in a fall (the subcontractor was fined $448,000).

If you think broadcasters are immune to such civil action, think again. A Mississippi TV production manager was electrocuted while setting up a live shot outside a local city hall. He had touched the side of the station's van, unaware that its boom antenna was touching an 8,000 V electrical wire. A few months earlier another station employee was electrocuted while working on the station's tower. OSHA's findings on these incidents had not been published when this chapter was submitted.

The *Florida Times-Union* reported a worker's death on October 14, 1997, in Jacksonville. The 28-year-old man was pronounced dead at the scene where he had been painting a billboard. He apparently had been electrocuted by nearby power lines and fell to the ground.

On October 23, 1997, the *Commercial Appeal Memphis,* TN reported that three men were killed when a 2,000 ft (.6 km) TV tower in Raymond, MI twisted and toppled to the ground. The workers were adjusting the guy wires when the accident occurred. One of the men was about 1,500 ft (.5 km) up when he radioed to others on the ground that "the tower is crumbling, get out of the way," just before falling to his death. OSHA investigations into this incident were anticipated as this book went to press.

Criminal Enforcement

In addition to civil penalties, employers are subject to criminal enforcement under certain circumstances. Here are a few examples of recent criminal actions:

On June 26, 1997, the owners of a South Carolina radio transmission company were sentenced to three months in prison for committing willful violations of OSHA. The incident in question involved the death of an employee who fell 150 ft (45.7 m) from a communications tower. The employee was not properly trained and was not provided with personal protective equipment. (U.S. v. Smith, DC MFla., No. 97-71-CR-J-21A).

In October, 1997, a Wisconsin jury convicted a malting company in the death of a worker who fell 100 ft (30.5 m) from a fire escape outside a grain elevator. The U. S. magistrate who imposed the fine of $450,000 following the guilty verdict cited "decades of indifference" on the part of the employer, and further stated that he wanted to send a message to "corporate America . . . that deliberate indifference to worker safety should come at high cost." At the time of sentencing, 31 other OSHA violations were pending in civil court against the company. (*Milwaukee Journal Sentinel,* December 18, 1996).

Case Law

On April 21, 1997, the *New Jersey Law Journal* reported that a former construction worker would receive $3.6 million from a settlement of his claim against Sea Isle City. The 32-year-old worker was rendered paraplegic in a fall. The complainant in the case alleged that the accident could have been prevented if there had been barriers, harnesses and other fall protection equipment.

The *Bureau of National Affairs, Boston,* reported on April 2, 1996 that five Boston firms agreed to pay $2.7 million to the family of a worker killed in the Boston Harbor cleanup project. The worker was crushed between two concrete sections being moved into place by a crane. In addition to the civil settlement with the four sub-contractors, a workers' compensation settlement was reached with the deceased's employer, the general contractor.

An OSHA official who investigated the case called the death "totally, absolutely, completely, and easily preventable." OSHA cited the employer, who paid penalties of $195,497, as well as the company that leased the crane, which paid $38,000 in penalties.

Controlling Employer Theory

OSHA uses a *controlling employer* or *multiple employer* theory to hold owners/operators liable for OSHA violations in addition to those imposed on the worker's actual employer. A prime example of this was reported by the *Bureau of National Affairs, Washington,* on May 12, 1997. The Occupational Safety and Health Review Commission (OSHRC) held an owner responsible for violations created by its independent contractor. The owner failed to ensure that its contractor followed lockout/tagout procedures during nightly cleaning of machinery at the owner's plant. The OSHRC found that the owner had supervisory authority and control over the work site and, although it was not the employer, the fact that it owned the hazardous equipment gave it responsibility to do what was "reasonably expected" to prevent the hazardous conduct. According to the OSHRC, "hazardous conduct by another employer's employee clearly is not beyond the reasonable control of all but the actual employer." (Secretary of Labor v. IBP Inc., OSHRC, No. 93-3059, 4/18/97).

It is important to note that, in addition to the OSHRC's ruling in this case and others there also exists in common law a theory of recovery against property owners for exposure of others to hazardous conditions on the property. The OSHRC's ruling and the common law theory pose a significant threat to broadcast owners/operators since liability is not limited to workers' compensation benefits as in the traditional employer-employee relationship.

9.7
DISASTER RECOVERY FOR BROADCAST FACILITIES

RICHARD RUDMAN
RADIO STATION KFWB, LOS ANGELES, CA

INTRODUCTION

Disaster planning and recovery is at once an art, a science and a technology. Disaster planners have their own professional groups and certification standards. If the resources are available, hiring a professional is a wise course of action. No matter who develops your disaster recovery strategy, an essential part of the process will be a written plan that will be periodically tested.

PREPAREDNESS 101

The private sector can learn from government emergency professionals, who recognize that *preparedness* is only the starting point of a continuous cycle.

The second step is the actual *response* to an emergency. *Recovery* comes third after immediate safety issues and other dangers are under control. The emphasis is restoring things to what passes for normal. The last phase is *mitigation* when the lessons of the emergency are studied. These lessons are applied as the circle is completed to another round of preparedness.

What Do You Prepare For?

Disaster planners think in terms of realistic risk. Simply put, it means do not spend as much time and effort planning for floods if the primary risk is earthquakes and the facility is on high ground. Those that have lived in an area for some time, will already have a fair idea of the greatest risks. If new to a region, research is a must. Be wary of advice from those who have lived there for awhile. They may overlook something that could be a significant risk. For instance, there are no 150-year-old transmitter sites. Long-term risks like 100- or 150-year floods may not have even been considered when the site was designed.

How Do You Prepare?

Disaster planners know that there are certain elements common to many types of preparedness. Both floods and earthquakes can isolate victims from utilities, supermarkets and emergency medical assistance.

Responsibility for Preparedness

Local emergencies can easily be triggered by preventable failures in air supply systems, roof leaks or uncoordinated telephone, computer or ac wiring changes. Seemingly harmless acts, such as employees plugging electric heaters into the wrong ac outlet, have brought down entire facilities.

Broadcast engineers should have responsibility (if not direct supervision of) the environmental infrastructure of broadcast facilities. This includes studios and transmitter sites. Without this critical element that tightly couples oversight to responsibility, critical electronic systems will be at the mercy of whomever or whatever controls the environment. Without responsibility, essentials such as proper installation practices, preventive maintenance, training and testing may fall by the wayside.

STARTING YOUR PLANNING PROCESS

Disaster recovery will be an easier task if some level of preparedness is in place. The process begins with a realistic risk assessment. The list should identify hazards based on local conditions.

Include the likelihood that threats from present or former employees who may hold grudges could turn into violent actions. Listeners and viewers sometimes get mad at broadcasters. Consultation with both the programming and security departments is strongly recommended. Some things to consider might be:

- Are there any nearby manmade hazards such as airports or chemical plants?
- What about nearby buried pipelines?
- What about the construction of the facility?[1]
- What hazardous materials are stored on the premises?
- Do not forget the obvious hazards like compressed gas cylinders and fuel for the emergency generator.

Risk Assessment and Business Survival

Once the list of realistic risks has been compiled, decide what to do to protect the staff and the facility. Preparedness depends on top management commit-

[1] Information from the Northridge, CA earthquake could rewrite seismic building codes for many types of structures. The Northridge quake showed that some high-rise structures thought to be quake-safe are not. Designers should be aware that seismic building codes usually allow for safe evacuation. They do not embody design criteria to prevent major structural damage. Quake-safe is not necessarily quake-proof.

ment. Without high-level support, funding and time to do proper preparedness will be hard to come by. Write a thoroughly researched report to top management in non-technical language. Relate risks to time off the air, potential for injury or death to staff or audience and potential damage to community image. Consider what 24 or more hours off the air would do to drive time ratings.

Adding Your Skills

Part of this process will rely on your broadcast engineering trouble-shooting skills and senses. Do not overlook the obvious. Computers, transmitters and telephone equipment depend on cool air. How long can they continue to operate during a heat wave if the one central building air system fails?

Take Practical Precautions

Preparedness should be considered in the same light as an insurance policy—the greater the risk, the higher the premium. Some communications facilities in earthquake country have three diesel generators and three air handling systems. Quake-resistant buildings sometimes rest on huge rubber isolators at the base of each supporting column.[2] These isolators can protect the structure and its contents from the most violent types of lateral movement during earthquakes. Proper seismic design and preparedness may be expensive or impossible for existing facilities, but may add no more than 10% to construction costs of a new building.

The Business Resumption Plan

A formal written *business resumption plan* (BRP) is just as important as an organization's written disaster plan. Some experts argue that the written disaster plan and the BRP should be formulated, tested and updated in concert. The prime directive: When disaster strikes, the first concern of an organization must be for the safety of its employees, customers, vendors and visitors.

The BRP is activated only when the basic requirements of life safety have been fulfilled. A BRP is a template to perform damage assessments, salvage operations, emergency restoration and, if needed, relocation of critical functions.

The focus of business resumption planning is maintaining or resuming core business activities following a disaster. The major goal of any BRP is resumption of production and delivery, customer service and notification and cash flow. In broadcasting, it begins with getting back on the air under adverse conditions when a facility is impaired.

What a BRP Can Do for You

A comprehensive and well designed BRP can accelerate recovery, thereby saving time, money and jobs. A good BRP is an insurance policy, if not a major investment. The following actions are the backbone of a BRP:

- Conduct a *business impact analysis* (BIA)[3]
- Promote employee buy-in and participation
- Seek input starting at the lowest staff levels
- Build a recovery strategy
- Build a validation process and testing procedure
- Assure a continuous update of the BRP.[4]

An Ounce of Prevention

While the title of this chapter incorporates the word *recovery*, the reader will not find a specific list of things to do that will magically prevent fatalities or injuries that have already taken place, or reconnect wiring to equipment lying on the floor in pieces. If faced with rehabilitating a facility where there has been no preparedness, competent broadcast engineers will eventually get the facility back on its feet. If you are faced with such a prospect, your main resource will be your ability to reach out to suppliers and other who you will need to begin the rebuilding process. Emergencies sometimes make this more difficult since others who have planned ahead may have written agreements in place with the very assets you may need on short notice. Roads may be impassible, limiting delivery possibilities. Recovery in such cases is not impossible; it just takes a lot longer. There is a lesson here you can convey to your owner or general manager.

Workplace Safety

Employers must always ensure safety in the workplace at all times. Some states like California have passed legislation that mandates that most employers identify potential hazards and protect their workers. Natural emergencies create special hazards that can maim or kill. A strong foundation of day-to-day safety can lessen the impact of major emergencies. For instance, assuring that plate glass in doors has a safety rating could avoid an accidental workplace injury, as well as compounding injuries during a major disaster.

Our Special Hazard

Special and dangerous hazards are found in the information and communications workplace. Tall equip-

[2] These rubber shock mounts are sometimes called base isolators. They are built using laminated layers of neoprene rubber and metal. The new Los Angeles County Emergency Operations Center is designed for 14 in. of lateral movement from normal or a bit more than 28 in. of total movement in one axis. The entire structure floats on 26 of these base isolators, designed to let the earth move beneath a building during an earthquake, damping transmission of rapid and damaging acceleration.

[3] A proper Business Impact Analysis (BIA) should yield the answer to the question of what level of damage your operation can sustain and still operate. When the cost of protecting against an emergency exceeds the benefit derived from protection, the law of diminishing returns sets in. It is far beyond the scope of this chapter to go into more detail on the BIA process.

[4] Thanks to Mary Carrido, President, MLC & Associates of Irvine, CA, for core elements of business resumption. It is far beyond the scope of this chapter to go into more detail on the BIA process.

ment racks are often not secured to floors, much less secured to load-bearing walls. Preventing equipment racks from tipping over during an earthquake may avoid crippling damage to both systems and people. Bookcases and equipment storage shelves should be secured to walls. Certain objects should be tethered, rather than firmly bolted. While securing heavy objects is mostly common sense, consult experts for special cases. Do not forget seismic rated safety chains for heavy objects like TV studio lights and large speakers.

Computers and monitors are usually not secured to work surfaces. A sudden drop from workstation height would ruin the day for most computers, video monitors and their users. Special Velcro® quick-release anchors and fasteners can support the entire weight of a personal computer or printer, even if the work surface falls over.

Bolting workstations to the floor and securing heavy equipment with properly rated fasteners can address major seismic safety issues. *G* forces[5] measured in an upper story of a high-rise building during the Northridge, CA, quake were greater than 2.7 times the force of gravity. A room full of unsecured workstations could do a fair imitation of a slam dance contest, even at lower accelerations. Cables can be pulled loose, monitors can implode and delicate electronics can be smashed into scrap. Even if your region has not been affected by seismic activity, consider that some parts of the earth have been given a long overdue rating by respected seismologists.

Glass Houses

Many local building codes now call for shatterproof glass for sliding doors and windows in homes and businesses. Large plate glass windows, including those used in studios, can become killers during major emergencies. A moment magnitude 7.4 earthquake can lethally hurl dagger-line shards from untreated plate glass windows through the air.[6] Extreme wind pressures during hurricanes can have a similar effect. The solution—purchase shatterproof glass or retrofit. Several companies manufacture a special film coating that can be applied to existing non-safety plate glass. This film will hold the glass together to reduce the danger of flying fragments.

Safety Maintenance

Maintaining safety standards is difficult in any size organization. A written safety manual outlining specific practices and procedures for normal workplace hazards as well as emergency-related hazards is not only a good idea, it may lower your insurance rates. If outside workers set foot in your facility, prepare a special safety manual for contractors. Include in it installation standards, compliance with *Lock-Out/Tag-Out*[7] and emergency contact names and phone numbers. Make sure outside contractors carry proper insurance and are qualified, licensed or certified to do the work for which they have been contracted.

RESCUE PLANNING FOR DIRE EMERGENCIES

When people are trapped and professionals cannot get through, our first instinct may be to attempt a rescue. More people are injured or killed in rescue attempts during major emergencies than are actually saved. Experts in Urban Search and Rescue (USAR) not only have the know-how to perform their work safely, they also have special tools that make this work possible under impossible conditions. The *Jaws of Life*, hydraulic cutters used to free victims from wrecked automobiles, is a common USAR tool. Pneumatic jacks can lift heavy structural members in destroyed buildings to free trapped people.

As a broadcast engineer you may never be faced with a life or death decision concerning a rescue when professionals are not available. However, those in the facilities you are responsible for may be faced with tough decisions. Consider that your planning could make their jobs easier or infinitely more difficult. Also consider recommending USAR training for those responsible for on-line management and operations of the facilities as a further means to ensure readiness.

The Recovery Process

The disaster has happened. Fatalities have been identified and the injured are being given first aid. What happens next?

Facility Integrity: Roof, Walls and Floor

A qualified person must first inspect the facility to make sure that it is safe. If you have just experienced the full effects of a hurricane or significant earthquake, damage may or may not be obvious. Most construction codes provide only that a structure must be built strong enough to allow evacuation without loss of life or

[5] One *G* is equal to the force of earth's gravity. An acceleration of 2 *Gs* doubles the effective force of a person or object in motion and nullifies the effectiveness of restraints that worked fine just before the earthquake. Force = (Mass) × (Acceleration). Seismic accelerations cause objects to make sudden stops—60 to zero in one second.

[6] The Richter Scale is no longer used by serious seismic researchers. Moment magnitude calculates energy release based on the surface area of the planes of two adjacent rock structures (an earthquake fault) and the distance these structures will move in relation to one another. Friction across the surface of the fault holds the rocks until enough stress builds to release energy. Some of the released energy travels via low frequency wave motion through the rock. These low frequency waves cause the shaking and sometimes violent accelerations that occur during an earthquake. For more on modern seismic research and risk, please refer to the bibliography for this chapter.

[7] *Lock-Out/Tag-Out* is a set of standard safety policies that assure that energy is removed from equipment during installation and maintenance. It ensures that every member of a work detail is clear before power is reapplied.

serious injury. You may be faced with the ultimate irony of a building that meets all local code provisions for an earthquake but may be rendered unsafe by the actual event.

Circumstances Beyond Your Control

Serious and obvious damage may result in a revocation of your building occupancy permit. You may not even be allowed in the facility to remove broadcast equipment. Do you have a contingency plan in place for an alternate studio site? A remote truck may be your only resort. Consider installing a locked access box outside the building where you may be able to connect to your microwave or phone company provided studio transmitter link (STL).

Basic Assumptions

Assuming your facility is intact or you have a contingency plan to get a signal on the air, what other considerations come into play during recovery? Your critical considerations can range employee morale to erecting an emergency antenna or getting your computer system working again.

Outside Plant Communications Links

Your facility may be operational, but failure of a wire, microwave or fiber communications link could be devastating. All outside plant links discussed in the following section presuppose proper installation. For wire and fiber, this means adequate *service loops* (coiled slack) so quake and wind stresses will not snap taut lines. It means that the telephone company has installed terminal equipment that will not fall over in a quake or be easily flooded out. A range of backup options is available.

Outside Plant Wire

Local telephone companies still use a lot of wire. If your facility is served only by wire on telephone poles or underground in flood prone areas, you may want what the telephone industry calls *alternate routing*. Alternate routing from your location to another central office (CO) may be very costly since the next nearest CO is rarely close. Ask to see a map of the proposed alternate route. If it is alternate only to the next block or duplicates your telephone pole or underground risk, the advantage you gain will be minimal.

Essential Service Designation

Most telephone companies can designate an *essential service*—a limited block of telephone numbers at a given location for lifeline communications. Lines so designated are usually found at hospitals and public safety headquarters. Contact your local phone company representative to see if your facility can qualifies. Many broadcasters who have close ties to local government emergency management should qualify easily.

Public Pay Telephones Are Essential

Inform your news departments that pay phones in public areas are at the top of the priority list for restoration by phone companies. This is not true of private pay phones or pay phones not located in public areas.

Dial Tone Trivia

Most broadcast engineers are aware of the basics of the telephone system in use in the United States. What they may not know is that the luxury of getting dial tone the instant they pick up a telephone instrument is not the norm in many parts of the world. The lesson: If you pick up an instrument properly connected to an outside line and detect side tone, you will eventually get a dial tone.[8] Jiggling the hook switch or hanging up will delay your access to an available dial tone. Patience is not just a virtue you need to teach those having to make emergency calls; it is a necessity.

Microwave Links

Wind and seismic activity can cause microwave dishes to go out of alignment. Quake-resistant towers and mounts can help prevent alignment failure, even for wind-related problems. Redundant systems should be considered part of your solution. A duplicate microwave system might lead to a false sense of security. Consider a non-microwave backup such as fiber for a primary microwave link. Smoke, heavy rain and snowstorms can cause enough path loss to disable otherwise sound wireless systems.

When a major link failure does occur, you ideally will have recorded ahead of time azimuth and elevation readings that will enable more rapid restoration. A stock of emergency restoration hardware, including brackets, bolts and short tower sections is a wise precaution. An alternate STL path, even telephone lines, can likewise be prudent insurance.

Fiber Optics Links

If you are not a fiber customer today, you will be tomorrow. Telephone companies will soon be joined by other providers (including cable) to seek your fiber business. You may be fortunate enough to be served by separate fiber vendors with separate fiber systems and routing to enhance reliability and uptime. Special installation techniques are essential to ensure fiber links will not be damaged by earth movement, subject to vandalism or vulnerable to single-point failure.[9]

---

[8] If you blow or talk into the telephone transmitter and hear yourself in the receiver, what you hear is called side tone. Telephone systems provide a small amount of feedback so users can judge their voice levels. Side tone depends on dc voltage on the line, an almost 100% sign that the line is connected all the way back to your first phone company Central Office. Connection beyond that point will of course depend on the integrity of the rest of the public switched network (PSN).

[9] Single point failure can occur in any system. Single point failure analysis and prevention is based on simple concepts: A chain is only as strong as its weakest link, but two chains, equally strong, may have the same weak link. The lesson may be make one chain much stronger, or use three chains of material with different stress properties.

Fiber should be installed underground in a sturdy plastic sheath called an interliner.[10] This sheath offers protection from sharp rocks or other forces that might cause a nick or break in the armor of the cable or actually sever one or more of the bundled fibers. Systems that only have aerial rights-of-way on utility poles for their fiber may not prove as reliable in some areas as underground fiber. Terminal equipment for fiber should be installed in quake-secure equipment racks away from flooding hazards. Fiber electronics should have a minimum of two parallel dc power supplies that are in turn paralleled with rechargeable battery backup.

Sonet® technology is a proven approach you should look for from your fiber vendor. This solution is based on topology that looks like a circle or ring. A fiber optics cable could be cut just like a wire or cable. A ring-like network will automatically provide a path in the other direction, away from the break. However, fiber installations that run through unsealed access points such as manholes can be easy targets for terrorism or vandalism.

Satellite

Ku- or C-Band satellite is a costly but effective way to link critical communications elements. C-Band has an added advantage over Ku during heavy rain or snowstorms. Liquid or frozen water or thick smoke can disrupt Ku-Band satellite transmission. A significant liability of satellite transmission for ultra-reliable facilities is the possibility that a storm or major fire could cause a deep fade, even for C-Band links. Another liability is short but deep semi-annual periods of *sun outage* when a link is lost while the sun is focused directly into a receive dish. While these periods are predictable and last for only a minute or two, there is nothing that can prevent their effect unless you have alternate service on another satellite with a different sun outage time or a terrestrial backup.

EMERGENCY POWER

Uninterruptible power supplies (UPS) are now common in the information workplace. From small UPS that plug into wall outlets at a personal computer work station to giant units that can power an entire facility, they all have one thing in common—batteries.

UPS Systems

UPS batteries must be tested regularly. Once exceeded, a UPS battery is nothing more than an expensive door stop. Allow the UPS to go on-line to test it. Some UPS test themselves automatically. Routinely pull the UPS ac plug out of the wall for a manual test. Some UPS applications require hours of power, while some only need several minutes.

While UPS provide emergency power when the ac mains go dead, many are programmed with another electronic agenda. They protect the plugged in devices from *bad power*. Many diesel generators in emergency service are not sized for the load they have to carry nor do they have proper power factor correction. Computers and other devices with switching power supplies can distort ac power waveforms. The result: power degraded to the point that a smart UPS will not trust it. The solution may be as easy as using a less educated UPS if the equipment being protected can tolerate it.

After a UPS comes on-line, it should go back to sleep after the emergency generator picks up the load and charges its batteries. If it senses the ac equivalent of poison, it stays on or cycles on and off. Its battery eventually runs down and it crashes. The best defense is to test your entire emergency power system under full load. If a UPS cycles on and off to the point that its batteries run down, you must find the cause. Consult your UPS manufacturer, service provider or service manual to find out if your UPS can be adjusted to be more tolerant. Some UPS cycling cannot be avoided with engine-based emergency power, especially if heavy loads like air conditioner compressors switch on and off-line.

Generator Testing

Technicians sometimes believe that starting an emergency generator without equipment load is an adequate weekly test. If your generator is diesel-driven, this may lead to *wet stacking*. Wet stacking occurs when a generator is run repeatedly with no load or a light load. When the generator is asked to come on-line to power a full equipment load, deposits that build up during no-load tests prevent it from developing full power under load. Use the load the diesel has to carry during an emergency to run the test. If that is not possible, obtain a resistive load bank to simulate a full load for an hour or two of hard running several times per year.

Fuel Considerations

Fuel stored in tanks gets old. Old fuel is unreliable. Gum and varnish can form and fuel begins to break down. Certain forms of algae can grow in diesel oil. Fuel additives can extend the storage period and prevent algae growth. A good filtering system and a planned program of cycling fuel through it can extend storage life dramatically. Individual fuel chemical composition, fuel conditioners and the age and type of storage tank all affect useful fuel life.

There are companies that will analyze fuel. If necessary, they can filter out dirt, water and debris that can rob the engine of power. The cost of additives and fuel filtering is nominal compared to the cost of new fuel plus hazardous material disposal charges for old fuel. Older fuel tanks can spring leaks that either introduce water into the fuel or introduce costly hazardous materials clean-up projects. The tank will be out of service while it is being replaced.

[10] Interliners are usually colored bright orange to make them stand out in trenches, manholes and other places where careless digging and prodding could spell disaster.

Generator Maintenance

Most experts recommend a generator health check every six months. Generators with engine block heaters put special stress on fittings and hoses. Vibration can loosen bolts, crack fittings and fatigue wires and connectors. If your application is critical, a second generator may give you a greater margin of safety. Your generator maintenance technician should take fuel and crankcase oil samples for testing at a qualified laboratory. The fuel report will let you know if your storage conditions are acceptable. The crankcase oil report might find microscopic metal particles, which are an early warning of a major failure.

Mission dictates need—need dictates reliability. If the design budget permits, a second or even third emergency generator is a realistic insurance policy. When you are designing a facility you know must never fail, consider redundant UPS wired in parallel.[11] During major overhauls and generator work, make sure you have a local source for reliable portable power. High-power diesel generators in wheels are commonly used to supply field power for events from rock concerts to movie shoots. Check your local phone directory for listings under *Generators—Electric or Emergency Lighting Equipment.* If you are installing a new diesel, remember that engines over a certain size may have to be licensed by your local air quality management district. Permits must be obtained to construct and store fuel in an underground tank.

AIR HANDLING SYSTEMS

Clean, cool, dry and pollutant-free air in generous quantities is critical for modern communications facilities. If you lease space in a high-rise, you may not have your own air system. Many building systems often have no backup, are not supervised nights and weekends and may have uncertain maintenance histories.

Your best protection is to get the exact terms for air conditioning detailed in your lease. You may wish to consider adding your own backup system, a costly but essential strategy if your building air supply is unreliable or has no backup. Several rental companies specialize in emergency portable industrial-strength air conditioning. An emergency contract for heating, ventilating and air conditioning (HVAC) that can be invoked with a phone call could save you hours or even days of downtime. Consider buying a portable HVAC unit if you are protecting a critical facility. Do not forget to have several hundred feet of inexpensive flexible ducting and a lot of duct tape available in advance.

Wherever cooling air comes from, there are times when you need to make sure the system can be forced to recirculate air within the building, temporarily becoming a closed system. Smoke or toxic fumes from a fire in the neighborhood can enter an open system. Toxic air could incapacitate your staff in seconds. With some advanced warning, forcing the air system to full recirculation could avoid or forestall calamity. It could buy enough time to arrange an orderly evacuation and transition to an alternate site.

[11] Consult the vendor for details on wiring needs for multiphase parallel UPS installations.

Emergency Actions

If no recirculation function is designed into the system, be prepared to shut it down for as long as possible. This may buy time to force the outside dampers closed or allow the outside source of polluted air to dissipate.

WATER HAZARDS

Water in the wrong place at the wrong time can be part of a larger emergency or be its own emergency. A simple mistake like locating a water heater where it can flood out electrical equipment can cause short circuits when it finally wears out and begins to leak. Unsecured water heaters can tear away from gas lines, possibly causing an explosion or fire during an earthquake.

Floods

Your facility may be located near a source of water that could flood you out. Many businesses are located in flood plains that see major storms once every 100 or 150 years. If you happen to be on watch at the wrong time of the century, you may wish that you had either located elsewhere or that you had stocked a very large supply of sandbags in advance. Remember to include any wet or dry pipe fire sprinkler systems as potential water hazards.

Water and electricity create a special hazard to you and your staff. Do not trust rubber gloves, rubber rainwear or other protective great. Always play it safe and disconnect power when working around flooded areas.

ELECTROMAGNETIC PULSE PROTECTION (EMP)

The EMP phenomenon associated with nuclear explosions can disable almost any component in a communications system. EMP energy can enter any component or system coupled to a wire or metal surface directly, capacitively or inductively. Some chemical weapons can produce EMP, but on a smaller scale.[12]

FEMA and EMP

The Federal Emergency Management Agency (FEMA) has been involved in EMP protection since 1970 and is the Federal agency charged with the overall direction of the EMP program. FEMA provides de-

[12] FEMA publishes a three-volume set of documents on EMP referenced in the bibliography, they cover the theoretical basis for EMP protection, protection applications and protection installation.

tailed guidance and, in some cases, direct assistance on EMP protection to critical communications facilities in the private sector. AM, FM and TV transmitter facilities that need EMP protection should discuss EMP protection tactics with a knowledgeable consultant before installing protection devices on radio frequency (RF) circuitry. EMP devices such as gas discharge tubes can fail in the presence of high RF voltage and disable the facilities they are supposed to protect.

ALTERNATE SITES

No matter how well you plan, something still could happen that will require you to abandon your facility for some period of time. Government emergency planners usually arrange for an alternate site for their Emergency Operations Center (EOC). Communications facilities can sign mutual aid agreements. Sometimes this is the only way to access telephone lines, satellite uplink equipment, microwave or fiber on short notice. If management shows reluctance to share, respectfully ask what they would do if their own facility is rendered useless.

Security

It is a fact of modern life that disasters caused by humans must now be considered in the planning and risk assessment process. Events ranging from terrorism to poor training can cause the mightiest organization to tumble. The World Trade Center and Oklahoma City bombings are a warning to us all. Your risk assessment might even prompt you to relocate if you are too close to ground zero.

Hostile Takeovers

Federal Communications Commission (FCC) Rules still state that licensees of broadcast facilities must protect their facilities from hostile takeover. Breaches in basic security have led to serious incidents at a number of places throughout the country. It has even happened at major-market TV stations. Since it is much easier to prevent such incidents than to recover from them, here are a few suggestions:

- Approve visits from former employees through their former supervisors
- Escort non-employees in critical areas
- Secure roof hatches from the inside and have alarm contacts on the hatch
- Check for legislation that may require a written safety and security plan.[13]
- Use video security and card key systems where warranted
- Keep fences in good repair, especially at unattended sites
- Install and test alarms at unattended sites
- Limit places bombs could be planted
- Consider limiting the number of outside windows or using bullet-resistant glass
- Plan a safe way to shut the facility down in case of invasion.

[13] California Senate Bill 198 mandates that California businesses with more than 100 employees write an industrial health and safety plan for each facility addressing workplace safety, hazardous materials spills, employee training and emergency response.

WORKPLACE AND HOME: HAND-IN-HAND PREPAREDNESS

A critical facility deprived of its staff will be paralyzed just as surely as if all the equipment suddenly disappeared. Employees may experience guilt if they are at work when a regional emergency strikes and they do not know what is happening at home.

Guilt and Bad Decisions

Encourage employees to prepare their homes, families and pets for the same types of risks the workplace will face. Emergency food and water and a supply of fresh batteries in the refrigerator are a start. Battery-powered radios and flashlights should be tested regularly. If employees or their families require special foods, prescription drugs, eyewear, oxygen, over-the-counter pharmaceuticals, sun block or bug repellent, remind them to have an adequate supply on hand to tide them over for a lengthy emergency.

Workplace and Home Basics

Heavy objects, like bookcases, should be secured to walls so they will not tip over. Secure or move objects mounted on walls over beds. Make sure someone in the home knows how to shut off natural gas. An extra long hose can help with emergency fire fighting or with draining flooded areas. Suggest family *hazard hunts*. Educate employees on what you are doing to make the workplace safe. The same hazards that can hurt, maim or kill in the workplace can do the same at home.

Contact Person

An excellent preparedness measure is to identify a distant business, relative or friend who can serve as the emergency message center. Employees may be able to call Aunt Tilly from work to find out if their family is safe and sound. Disasters that impair telephone communications teach us that it is often possible to make and receive long distance calls when a call across the street will not get through. Business emergency planners should not overlook this hint. A location in another city or a key customer or supplier may make a good out-of-area emergency contact.

On the Road

Personal and company vehicles should all have emergency kits that contain basic home or business emergency supplies. Food, water, comfortable shoes and old clothes should be added. If their families are prepared at home or on the road, employees will have added peace of mind. It may sustain them until they can get home safely.

Expectations, 911 and Emergencies

Television shows depicting 911 saving lives over the telephone are truly inspirational. But during a major emergency, resources normally available to 911, including their very telephone system, may be unavailable. Emergency experts used to encourage people to be prepared to be self-sufficient at the neighborhood and business levels for 72 hours or more—some now suggest a week or longer.

Managing Fear

Normal human reactions when an emergency hits are colored by a number of factors, including fear. As the emergency unfolds, we progress from fear of the unknown to fear of the known. While preparedness, practice and experience may help keep fear in check, admitting fear and the normal human response to fear can help us keep our cool.

Some people prepare mentally by reviewing their behavior during personal, corporate or natural emergencies. They then consider how they could have been better prepared to transition from normal human reactions like shock, denial and panic to abnormal reactions like acceptance and steady performance. The latter behaviors reassure those around them and encourage an effective emergency team.

Other "Normal" Reactions

Most people experience a rapid change of focus toward their own personal well-being. "Am I ok?" is a very normal question at such times. Even the most altruistic people have moments during calamities when they regress. Even the best of us can temporarily become selfish. Once people realize they do not require immediate medical assistance, they can usually start to focus again on others and on the organization.

BIBLIOGRAPHY

Baylus, Ellen, *Disaster Recovery Handbook*, Chantico Publishing Company.

Federal Emergency Management Agency (FEMA), *Electromagnetic Pulse Protection Guidance*, Vols. 1–3, Washington, DC.

Fletcher, Robert, *Federal Response Plan*, Federal Emergency Management Agency, Washington, DC.

Handmer, John, and Parker, Dennis, *Hazard Management and Emergency Planning*, James and James Science Publishers.

Rothstein Associates, *The Rothstein Catalog on Disaster Recovery and Business Resumption Planning*, Rothstein Associates.

FOR FURTHER INFORMATION

Associations/Groups

Association of Contingency Planners, 14775 Ventura Boulevard, Suite 1-885, Sherman Oaks, CA 91483.

Business and Industry Council for Emergency Planning and Preparedness (BICEPP), PO Box 1020, Northridge, CA 91328.

Disaster Recovery Institute, 1810 Craig Road, Suite 125, St. Louis, MO 63146. (DRI holds national conferences and publishes the *Disaster Recovery Journal*.)

Earthquake Engineering Research Institute (EERI), 6431 Fairmont Avenue, Suite 7, El Cerritos, CA 94530.

National American Red Cross, 2025 E Street, NW, Washington, DC 20006.

National Center for Earthquake Engineering Research, State University of New York at Buffalo, Science and Engineering Library-304, Capon Hall, Buffalo, NY 14260.

National Coordination Council on Emergency Management (NCCEM), 7297 Lee Highway, Falls Church, VA 22042.

National Hazards Research and Applications Information Center, Campus Box 482, University of Colorado, Boulder, CO 80309.

Business Recovery Planning

Harris Devlin Associates, 1285 Drummers Lane, Wayne, PA 19087.

Industrial Risk Insurers (IRI), 85 Woodland Street, Hartford, CT 06102.

MLC & Associates, Mary Carrido, President, 15398 Eiffel Circle, Irvine, CA 92714.

Price Waterhouse Coopers, Dispute Analysis and Corporate Recovery Dept., 555 California Street, Suite 3130, San Francisco, CA 94104.

Resource Referral Service, PO Box 2208, Arlington, VA 22202.

The Workman Group, Janet Gorman, President, PO Box 94236, Pasadena, CA 91109.

Life Safety/Disaster Response

Caroline Pratt & Associates, 24104 Village #14, Camarillo, CA 93013.

Industry Training Associates, 3363 Wrightwood Drive, Suite 100, Studio City, CA 91604.

Emergency Supplies

BEST Power Technology, PO Box 280, Necedah, WI 54646 (UPS).

Exide Electronics Group, Inc., 8521 Six Forks Road, Raleigh, NC 27615.

Extend-A-Life, Inc., 1010 South Arroyo, Parkway #7, Pasadena, CA 91105.

Velcro® USA, PO Box 2422, Capistrano Beach, CA 92624.

Worksafe Technologies, 25133 Avenue Tibbets, Building F, Valencia, CA.

Construction/Design/Seismic Bracing

American Institute of Architects, 1735 New York Avenue, NW, Washington, DC 20006.

DATA Clean Corporation (800-328-2256).

Geotechnical/Environmental Consultants

H. J. Degenkolb Associates, Engineers, 350 Sansome Street, San Francisco, CA 94104.

Leighton and Associates, Inc., 17781 Capen, Irvine, CA 92714.

Miscellaneous

Data Processing Security, Inc., 200 East Loop 820, Fort Worth, TX 76112.

EDP Security, 7 Beaver Brook Road, Littleton, MA 01460.

ENDUR-ALL Glass Coatings, Inc., 23018 Ventura Blvd., Suite 101, Woodland Hills, CA 91464.

Mobile Home Safety Products, 28165 B Front Street, Suite 121, Temecula, CA 92390.

Commercial Filtering, Inc., 5205 Buffalo Avenue, Sherman Oaks, CA 91423 (Fuel Filtering).

INDEX

NAB